DTIC FILE COPY

1

ARI Research Note 88-35

AD-A196 727

TRAINING EFFECTIVENESS AND COST ITERATIVE TECHNIQUE (TECIT) VOLUME I: TRAINING EFFECTIVENESS ANALYSIS

Isadore Goldberg
The Consortium of Washington Area Universities

for

Contracting Officer's Representative
John J. Kessler

Training and Simulation Technical Area
Robert J. Seidel, Chief

TRAINING RESEARCH LABORATORY
Jack H. Hiller, Director

U. S. Army

Research Institute for the Behavioral and Social Sciences

April 1988

Approved for public release; distribution unlimited.

U. S. ARMY RESEARCH INSTITUTE

FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the

Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

WM. DARRYL HENDERSON
COL, IN
Commanding

Research accomplished under contract
for the Department of the Army

The Consortium of Washington Area Universities

Technical review by

Douglas Dressel

8 UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

ADA196727

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION: Unclassified	1b. RESTRICTIVE MARKINGS: n/a
2a. SECURITY CLASSIFICATION AUTHORITY: n/a	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE: n/a	
4. PERFORMING ORGANIZATION REPORT NUMBER(S): n/a	5. MONITORING ORGANIZATION REPORT NUMBER(S): ARI Research Note 88-35
6a. NAME OF PERFORMING ORGANIZATION: Consortium of Washington Area Universities; 6b. OFFICE SYMBOL (If applicable): - -	7a. NAME OF MONITORING ORGANIZATION: U.S. Army Research Institute for the Behavioral and Social Sciences
6c. ADDRESS (City, State, and ZIP Code): 1717 Massachusetts Ave., NW, Suite 101, Washington, DC 20036	7b. ADDRESS (City, State, and ZIP Code): 5001 Eisenhower Avenue, Alexandria, VA 22333-5600
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: n/a; 8b. OFFICE SYMBOL (If applicable):	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: MDA903-82-C-0383
8c. ADDRESS (City, State, and ZIP Code): n/a	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO	WORK UNIT ACCESSION NO
6.37.44.A	2Q263744A795	3.4.4	3.4.4.C.1

11. TITLE (Include Security Classification)
Training Effectiveness and Cost Iterative Technique (TECIT)
Volume I: Training Effectiveness Analysis

12. PERSONAL AUTHOR(S) Isadore Goldberg

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final Report	FROM Feb.86 TO Feb.87	April 1988	185

16. SUPPLEMENTARY NOTATION Dr. Goldberg is on the Psychological Team at the University of the District of Columbia. Volume II "Cost Effectiveness Analysis" is authored by (OVER)

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Computer Assisted Instruction (CAI), Cost and Training Effectiveness Analysis (CTEA) ~~models~~, Training Effectiveness Analysis (TEA) ~~models~~, (OVER)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This research note describes the effectiveness model of the Training Effectiveness and Cost Iterative Technique (TECIT), a new model concerned with the cost effectiveness of training devices and simulators (TD/S) at all phases of the life cycle development. Volume II describes the cost model.

For the purposes of this study, the effectiveness of a training device or simulator was defined as a function of the following factors: safety, acquisition learning on the TD/S, transfer of training from the TD/S to an exercise on the weapon system during training, job or battle readiness, and the utilization ratio of the TD/S.

A research strategy is outlined in the research note. This strategy considers cross-sectional and longitudinal designs, TD/S life cycle phases, and various validity designs (e.g. discriminant, concurrent, and predictive validity). Sampling of subject matter experts' opinions and TD/S is also considered.

(OVER)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION: Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL: John J. Kessler	22b. TELEPHONE (Include Area Code): 202/274-8694; 22c. OFFICE SYMBOL: PERI-II

DD Form 1473, JUN 86 *Previous editions are obsolete.* SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

ARI RESEARCH NOTE 88-35

16. Supplementary Notation (continued)

A.V. Adams and M.L. Rayhawk of the George Washington University.
John J. Kessler, contracting officer's representative.

18. Subject Terms (continued)

Training Devices,
Simulators,

19. Abstract (continued)

Also included in this document is a review of related models, including the Device Effectiveness Forecasting Technique (DEFT), Forecasting Training Effectiveness (FORTE), and Comparison Based Prediction (CBP). A comparison of model features is also included, along with sample questionnaires, and an illustrative data base.

ACKNOWLEDGMENTS

This report was prepared with support from the Consortium of Washington Area Universities and the Army Research Institute for the Behavioral and Social Sciences. We are indebted to a number of individuals who have made important contributions to the study. We wish to thank the Army Research Institute Advisory Committee, including D. Bruce Bell, Stanley F. Bolin, Hyder Lakhani, Donald Haggard, Harold Wagner and John J. Kessler, who offered advice and counsel throughout the study. Angello Mirabella and Jesse Orlansky are accorded a special note of appreciation for their detailed critique of the draft of this report and their many useful insights. A special note of appreciation is offered to John J. Kessler who served as the Contracting Officer's Representative. Dr. Kessler provided valuable assistance throughout the effort serving as a sounding board for our ideas and making many useful suggestions for improving the final report.

During the course of the study, several individuals at the Ft. Knox Armor Center provided us with access to specialized literature and insights into the conduct of training using computer-assisted instruction, training devices and simulators. These insights have materially enriched the study. Among others who contributed in this respect, we want to especially thank Donald F. Haggard, Donald M. Kristiansen and John A. Boldovici. From our own staff we wish to thank Nidhi Khattri for invaluable research assistance and Twannah Ellington for typing and production and Helen Young for editing of this report. As is customary, the views expressed in this study are our own and do not necessarily reflect the official views of the U.S. Army. We naturally assume responsibility for any errors or omissions.

BRIEF

This document describes the effectiveness submodel of TECIT, a new multipurpose model concerned with the cost effectiveness of training devices and simulators (TD/S) at all phases of life cycle development. Volume II describes the cost model.

The effectiveness of a training device or simulator is defined as a function of the following: safety; acquisition learning on the TD/S; transfer of training from the TD/S to an exercise on the Weapon System (WS) during training; job or battle readiness (alternately defined as the transfer of training from the TD/S to the job, a battle exercise after training or the retraining schedule needed to maintain readiness); and the utilization ratio of the TD/S. The analyst selects those elements appropriate to the TD/S in question. Applications of the model and research are given equal attention.

The training effectiveness submodel has two components: (1) Problem Analysis and Definition Component; (2) Analytic Component. The Problem Analysis and Definition Component guides the analyst in considering and documenting items such as the following: (a) the application(s) for which the analysis is being made (i.e., concept, development, fielding or research, system vs. non-system training, single course or multi-course applications, personnel to be trained, Weapon system(s) and course(s) to which the TD/S is applicable, and placement of the TD/S in the course and the career sequence); (b) life cycle development phases of the WS(s), training program(s) and TD/S; and (c) study team and SME characteristics - roles, responsibilities, background, experience and effort expended.

This component also guides the gathering of information about the WS(s), the training program(s), the TD/S, predecessor TD/S, similar TD/S and databases relevant to the application(s) to be made. It aids analysts in making preliminary estimates of TD/S effectiveness and in providing information to Subject Matter Experts (SMEs) for making analytic judgments. It also aids in identifying appropriate SMEs and documents an audit trail of information for further applications and research. A task/subtask/skill comparison method aids in comparing baseline (predecessor or similar) TD/S with the proposed TD/S for initial design or improvement.

In the Analytic Component, the analyst makes estimates of each appropriate effectiveness element or obtains them from Subject Matter Experts (SMEs). The method employed, judgmental variance estimating, enables quantitative estimates to be made of important sources of variance that may affect the design of the TD/S. Examples of judgmental

variance sources include those attributable to trainees, tasks, the criterion, team training, physical fidelity, functional fidelity, and instructional management. Estimates may be made at a task level or for the TD/S as a whole.

Time and performance measures of acquisition learning and transfer of training are used for in-course measures. Rating scales and checklists are used for post-course transfer, safety and instructional management. These methods lend themselves to obtaining quantitative measures of reliability and validity.

In early phases of the TD/S life cycle, analytic methods are employed (bolstered by databases, predecessor TD/S and similar TD/S) to conceptualize, design and develop the TD/S. No empirical data are available on the new TD/S. In the fielding phase, attention turns to analytic methods to support empirical studies of acquisition, transfer and utilization. TECIT provides a number of quantitative methods of organizing judgmental data to forecast or support empirical data.

Since the analytic judgmental methods yield quantitative estimates of variability, reliability and validity, the model may be used for both research and applications. The central research issues are: (1) What is the accuracy of analytic estimates? (2) What methods and aids can be employed by analysts to make them more accurate? (3) To what extent, under what circumstances, and for what applications are analytic estimates a useful complement to empirical data? (4) To what extent and for what applications can analytic estimates serve as proxies for empirical data?

A research strategy is outlined. The research strategy considers cross-sectional and longitudinal designs, TD/S life cycle phases, and various validity designs (i.e., discriminant, concurrent, and predictive validity). Sampling of SMEs and TD/S is also considered. Accuracy of prediction is considered the most important characteristic of validity designs.

Using data available from the audit trail, research can also be conducted to assess the effort and cost required to exercise the model under various conditions of availability of information and life cycle phases of the weapon system and training program. A validation plan is presented for testing the model on the Tank Commander's Basic Non-Commissioned Officers Course for the M1 Abrams Tank at the Ft. Knox, Kentucky Armor School.

A review of related models is included in this document including the Device Effectiveness Forecasting Technique (DEFT), Forecasting Training Effectiveness (FORTE), and Comparison Based Prediction (CBP). A comparison of model

features is also included, along with sample questionnaires and an illustrative data base.

Recommendations for further development of TECIT include the development of a user's guide, a research guide, and computerization of the model.

TECIT was designed for use by the Training Technology Field Activity (TTFA), the Army Research Institute and the Armor School at Ft. Knox, Kentucky. However, it should be useful to all military personnel and contractors concerned with the design and development of TD/S and to researchers interested in the improvement of analytic and empirical methods aiding this process.

TABLE OF CONTENTS

TABLES

FIGURES

FORMS

Chapter 1

OVERVIEW OF THE TRAINING EFFECTIVENESS AND COST ITERATIVE TECHNIQUE (TECIT)

INTRODUCTION

This report (in two volumes) is part of a long-range Army Research Institute program to develop more efficient models and methods for assessing training device and simulator cost effectiveness. The report explains the rationale for a new model, details its features, and outlines a plan to validate it. Volume I covers the entire model but focuses on training effectiveness analysis. Volume II details the rationale and methodology for cost analysis. Together they combine training measures with cost analysis in a model which can help the Army choose the type and amount of training needed to "produce" different kinds of skills. They do so by building upon and integrating some of the best features of many past research efforts (e.g., Adams and Rayhawk, 1986; Goldberg and Khattri, 1986; Klein, 1985; Knerr et al., 1985; Orlansky and String, 1977, 1979, 1981; Orlansky, 1985; Pfeiffer et al., 1985; Pfeiffer and Scott, 1985; Rose and Wheaton, 1984).

RESEARCH OBJECTIVE

Produce a model of cost and training effectiveness analysis (CTEA) which can be applied to the development of training devices and simulators (TD/S) at all phases of the TD/S and weapon system acquisition process.

BACKGROUND

Many models show how to plan and analyze training programs in the conceptual phase of new weapon system acquisition (e.g., HARDMAN, Training Effectiveness Cost Effectiveness Prediction), but few deal specifically with TD/S development. Those that do, do not apply to all stages of TD/S acquisition and do not address costs (Adams and Rayhawk, 1986; Goldberg and Khattri, 1986). The model described in this report is designed to overcome these and a number of other deficiencies in CTEA.

TECIT was particularly developed to serve the needs of Training Tecnnology Field Activities (TTFA), newly formed efforts within the Training and Doctrine Command (TRADOC) charged with the improvement of training in general, technology transfer, and the exportability of "packages" of military training. To serve these needs the TECIT model has been designed for application to TD/S concept and design phases as well as to issues of exportability and technology transfer.

ORGANIZATION OF THE REPORT

This chapter presents an overview of the TECIT Training Effectiveness Submodel and a review of related models. Chapter 2 presents the Problem Definition and Analysis Component of training effectiveness, while Chapter 3 presents the Analytic Component of training effectiveness and summarizes a cost analysis method. Chapter 4 presents a research strategy and validation plan. Volume II details the costing method and the integration of costs and effectiveness.

SUMMARY OF TECIT CHARACTERISTICS

General Approach

The TECIT model incorporates other models within it, e.g., Device Effectiveness Forecasting Technique, Forecasting Training Effectiveness, and Comparison Based Prediction. However, TECIT combines criterion measures many of which have not been included in past indices of training effectiveness, namely safety and emergency procedures, job readiness for a work sample TD/S, and utilization. Transfer of training within a course is the one paradigm which uses the empirical transfer experiment and for which the other models appear to have been developed.

Both analytic and empirical tests of TD/S effectiveness may be employed depending on the phase of development of the TD/S. For example, in the conceptual and design phases of a TD/S, only analytic methods can be used, supported in some cases by databases or comparison cases from other TD/S. Although databases and comparison cases are useful, they do not provide empirical data on the new TD/S. No empirical data can be obtained since the new TD/S has not yet been developed. After a TD/S has been fielded, the accumulation of empirical data specific to that TD/S becomes a primary concern. However, because of many practical and research design constraints, the empirical data and methods for measuring the effectiveness of a TD/S are often limited. Thus some means is needed to effectively employ both analytic and empirical methods as a TD/S evolves through its life cycle. The relative emphasis on analytic methods vs. empirical methods shifts, depending on whether the TD/S is in the conceptual phase or whether it has been fielded, however, both analytic and empirical methods are potentially useful at all phases of TD/S development.

The applications that a model needs to address also differ in the conceptual vs. the fielding phases. The conceptual phase is concerned with issues such as deciding whether or not a TD/S is needed, evaluating alternative design concepts and guiding the development process. All of these applications require analytic methods. Real

alternative TD/S are rarely, if ever, developed for empirical testing. After fielding, emphasis shifts to implementation, installation, deployment, technology transfer, demonstrating effectiveness of the TD/S and exportability, processes which take many years. It is these processes that lend themselves to obtaining empirical tests of device effectiveness.

TECIT is also designed to aid in problem definition and analysis and to obtain analytic estimates of appropriate variables in a form that facilitates research and validation.

The research approach emphasizes accuracy of analytic estimates. The accumulation of empirical data may be used as criteria for measures for analytic methods in longitudinal studies. Alternately, cross-sectional research studies or studies on databases may attempt to establish how well TECIT measures discriminate among various TD/S characteristics.

Definitions of TD/S

For the purpose of formulating the model TD/S are defined in terms of their functions and purposes as follows (adapted in part from Blaiwes and Regan, 1986):

1. TD/S are those technologies oriented primarily to learning, integrating, and practicing job performance skills in a physical and learning environment that simulates the job skills in question. The TD/S incorporates a degree of similarity to the real world environment that is greater than training technologies and delivery systems ordinarily employed in a conventional classroom environment and enable skills to be exercised in a manner conducive to learning.

2. Work sample or criterion-referenced TD/S are those that are able to represent job or battle conditions that would be infrequently encountered on the job, may be life threatening, and for reasons of time and costs could not otherwise be included in training. These TD/S are expected to improve job or battle readiness. Examples include maintenance simulators that represent a wide array of breakdowns and tactical and strategic simulators that prepare trainees for a broad array of battle conditions.

3. Safety. Some TD/S are designed to provide a safe learning environment. There is evidence to show that simulator experience helps reduce accidents.

4. "Training considerations generally favor simulators.

Foremost among these are mechanical reliability, availability of training time, compression and re-arrangements of training sequences, and freedom from limiting factors (e.g., weather, air congestion)." (Blaiwes and Regan, 1986).

5. Costs. As a practical matter, there is usually a higher magnitude of investment (or research and development) cost associated with developing TD/S as opposed to training aids for conventional classroom instruction.

   However, comparing TD/S and WS, Blaiwes and Regan (1986) point out: "cost differentials between simulators and job equipment in construction, utilization, and amortization are generally significantly in favor of the simulator when it is used efficiently in conjunction with the actual equipment, classroom instruction and the like."

Thus, important distinctions between a TD/S and classroom instruction lie in their realistic representation of performance skills as opposed to knowledge and information, the opportunity to integrate knowledge and skills in a realistic environment, and relative costs. Important distinctions between learning on the TD/S and the WS lie in work sampling, safety, cost advantages and training advantages.

From a modeling standpoint, it is also important to distinguish between the TD/S hardware, software and courseware. In many cases, a TD/S hardware configuration may be considered as a carrier of software and courseware, so that part of the design goal is to develop hardware with sufficient flexibility to be used with a variety of software and courseware. Multi-course TD/S must therefore be distinguished from single course TD/S. The term TD/S will refer in this report to the software and courseware applicable to a single specific course of instruction but potentially exportable to a number of settings.

These definitions and distinctions between TD/S and conventional instruction and TD/S and learning on the job or WS itself are useful as guides to measurement of TD/S outcomes. Other definitions are given in Appendix C.

The Structure of TECIT

TECIT is composed of two submodels:

1. TD/S effectiveness submodel (this volume)
2. TD/S life cycle costs submodel (volume II)

The TD/S effectiveness submodel is composed of two components as shown in Figure 1.

1. The Problem Definition and Analysis Component (Chapter 2)

2. The Analytic Component (Chapter 3)

Problem Definition and Analysis Component

The problem definition and analysis component guides the analyst to consider and document the following. Eight forms are used.

1. Training Spectrum Analysis - defines system vs. non-system training, single course or multi-course applications, personnel to be trained, weapon system(s) and course(s) to which the TD/S is applicable, and placement of the TD/S in the course and the career sequence.

2. Life Cycle Development Phases of the WS(s) and training program(s) are indicated.

3. Life Cycle Phase of the TD/S and Purposes of the Analysis - selects and documents the application(s) for which the analysis is being made, i.e., concept development, fielding, exportability or research.

4. Information Gathering - guides the gathering of information about the WS(s), the training program(s), the TD/S, predecessor TD/S, similar TD/S and databases relevant to the application(s) to be made: an aid to making preliminary estimates of TD/S effectiveness, to providing information to SMEs for making analytic judgments; and an aid to identifying appropriate SMEs and documenting an audit trail of information for further applications and research. An illustrative database is given in Appendix A.

5. Task/subtask/skill comparison - an aid for comparing baseline (predecessor or similar) TD/S with the proposed TD/S for initial design or improvement; an aid to judgments about task similarity and the relative weight to be given to baseline data and the new threat scenario; an aid for comparing training program, TD/S and tasks to judge which tasks need to be taught in each.

6. Baseline Data Analysis-summary of data obtained from 4 and 5 above.

7. Documenting Study Team and SME Characteristics - an aid for guiding and documenting roles, res-

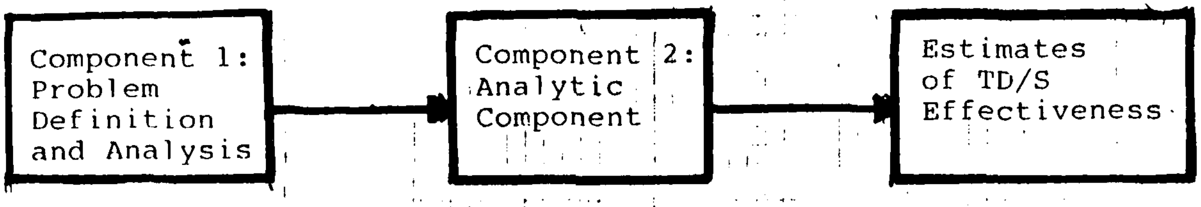

Figure 1. Flowchart of the TECIT Training Effectiveness Submodel

ponsibilities, background, experience and effort expended; a research tool for comparing judgments by background and experience.

8. Is a TD/S needed?-a brief checklist for a preliminary determination of this issue.

The forms can be used manually or, with further development, would be contained on a computer.

Analytic Component

The analytic component is made up of two parts:

Part 1: TD/S effectiveness is defined as a function of acquisition learning on the TD/S, transfer of training in the course, safety (accident reduction), job readiness, and the utilization ratio.

Part 2: Judgmental variance sources and instrument file. Identifies sources of variance and instruments for estimating effectiveness elements.

Part 1 - TD/S Effectiveness function:
The effectiveness of a training device or simulator is defined as a combination of the following: acquisition learning on the TD/S; safety or accident reduction; transfer of training from the TD/S to an exercise on the weapon system (WS) during training; job (or battle) readiness; and the utilization ratio of the TD/S.

This function* may be written as follows:

$$\text{TD/S E (f)} = \left\{ \frac{\text{S, ToT, JR}}{\text{Acq}} \right\} \text{UR}$$

Where

TD/S E refers to the training effectiveness function.

Acq. is acquisition learning on the TD/S measured in terms of time to criterion

*J. Orlansky does not agree with this function. Letter of 11/26/86.

on the TD/S

S is a safety rating

ToT is transfer of training from the TD/S to an exercise on the WS during training measured in various ways such as time savings or performance gains on the WS attributable to training on the TD/S.

JR is a rating of job readiness for a work sample TD/S, alternately defined as the transfer of training from the TD/S to the job, transfer to a battle exercise after training, or the skill maintenance retraining schedule required to maintain readiness.

UR is the utilization ratio of the TD/S defined as the hours used divided by the hours scheduled times 100.

It should be noted that three elements in the formula, namely, safety, transfer of training within the course, and job readiness may each be relevant to different TD/S. If not relevant, their values reduce to zero. The Multi-Attribute Utility Assessment Method (MAUM) method is used to combine the various elements as each element is not expressed in comparable metrics. The MAUM method allows the analyst to combine results of the elements according to their criticality and importance. The combination of the three elements is divided by acquisition (Acq) time to criterion to reflect an efficiency ratio of transfer to acquisition. The utilization rate multiplier reflects the idea that effectiveness will not be achieved unless it is used as scheduled. The effectiveness function is used in relation to costs in a cost-effectiveness analysis.

As given, this function is most useful in the concept and design phase of TD/S development to compare two or more concepts and select among them. Elements of the function are used in the fielding phase of TD/S development.

Acquisition, transfer of training within the course, and the utilization ratio are expressed as metrics identical to their empirical measurement in training to enable comparisons to be made of analytic and empirical estimates for validation purposes. Safety and job readiness use ratings because quantitative expression would be too demanding for analysts to determine, empirical data will not be available for comparison purposes for many years after the TD/S is conceived, and these measures can only be indirectly validated.

Acquisition on the TD/S is a central element in that judgments about safety, transfer of training and job

readiness all impact time, performance and the criterion in TD/S acquisition. For example, if safety is a concern, there must be sufficient practice on the TD/S to assure that the trainee is ready to practice on the WS. The ToT paradigm is of interest only for safe tasks. Similarly, if a work sample TD/S is designed, there must be sufficient practice on the TD/S to assure job or battle readiness or minimize the retraining schedule. The transfer of training to a WS exercise within the course may or may not be of interest.

The analyst selects those elements of interest appropriate to the TD/S in question. Elements not relevant to a particular application such as safety or job readiness are reduced to zero and ignored by the analyst. Safety, acquisition, transfer, and job readiness analyses may be conducted at the task level as well as for the TD/S as a whole. The utilization rate analysis is conducted for the TD/S as a whole; a task level analysis is not conducted.

For in-course transfer the decision required of the analyst in selecting appropriate data and formula elements is whether to select time to criterion (time variable, performance fixed) measures or performance measures (performance variable, time fixed) as the primary method of analysis. This decision is based on how training on the WS is structured and may be discerned by examining the WS exercise and information gathered in Component 1. The formula should also correspond to that selected for acquisition.

An illustration of data entry and calculations proceeds as follows:

1. Primary measure is time to criterion. The analyst enters estimates of the following data items:

   1.1 WS - time to criterion on the WS for the control group.

   1.2 WS(TD/S) - time to criterion on the WS for the transfer group.

   1.3 TD/S - time to criterion on the TD/S for the transfer group.

   The following summary measures are then calculated:

   1.4 Transfer Effectiveness Ratio =

$$\frac{WS - WS(TD/S)}{TD/S}$$

1.5 Percent Time Saved (PTS) on the WS

$$= \frac{WS - WS(TD/S)}{WS} \times 100$$

1.6 Proportion Total Training Time Saved/Added (PTTS/A)

$$=1 + \frac{[WS(TD/S) + TD/S] - WS}{WS}$$

If a secondary measure of performance is desired in addition to the time to criterion measure, the analyst enters the estimates of the following items of interest:

1.7 The criterion (Crit.) value of the performance measure.

1.8 Transfer (T) group performance average on the WS.

1.9 Control (C) group performance average on the WS.

The following summary measure is then calculated:

Percent Transfer to Criterion (PTC) =

$$\frac{T}{\text{Crit.}} \times 100 - \frac{C}{\text{Crit.}} = \frac{T - C}{\text{Crit.}} \times 100$$

2. Primary measure is performance. The analyst starts by entering estimates of the following data items:

2.1 Transfer (T) group average on the WS.

2.2 Control (C) group average on the WS.

2.3 Scale direction: High score means better performance or low score means better performance.

Depending on information available, the following is also entered:

2.4 The Criterion (Crit.) value of the performance measure (e.g., the combat performance standard).

2.5 The maximum score of the performance measure when a high score means better performance.

One or more of the following summary measures of performance transfer are then selected and calculated

depending on the information available in 2.4 and 2.5 and the analyst's interests:

2.6 Percent Transfer to Criterion (PTC) =

$$\frac{T}{Crit.} \times 100 - \frac{C}{Crit.} \times 100 = \frac{T - C}{Crit.} \times 100$$

when the criterion value is available.

2.7 Percent Transfer Max. (PTM) =

$$\frac{T - C}{Max} \times 100$$

when there is a maximum score.

2.7 Percent Transfer $= \frac{T - C}{T + C} \times 100$

when the criterion value has not been specified and there is no maximum score.

There are differences in the usefulness and interpretation of these formulae when a low score represents better performance. See the Technical Discussion of Performance Measures in Chapter 3. A computer routine would take these variations into account.

Time estimates used when performance is of primary interest are "fixed" times for acquisition on the TD/S and the WS for both groups to reach the performance level indicated. The PTTS/A formula (using fixed times) is calculated to determine the impact of adding the TD/S on total training time. These points are illustrated in Chapter 3.

In practice, the data profile would be obtained at a task level for diagnostic analyses as well as for the TD/S as a whole. When the tasks for two alternative TD/S concepts differ in some respects, task level analyses are required to avoid distorted inferences. Task level analyses are also required when the comparison is between tasks or skills that might be taught by "conventional" instruction vs. the TD/S. In the design phase, "what if" questions can be posed regarding physical and functional fidelity and trade-offs among acquisition, transfer, performance, time, accident reduction and costs. The analyst may use all or part of the data elements appropriate to the problem.

It can be noted that all data elements are considered along with traditional summary measures of transfer of training such as the Transfer Effectiveness Ratio (TER), the

Percent Time Saved on the WS, and various Performance Percent Transfer measures. The limitations of these summary measures are discussed in Chapter 3. The Multi-Attribute Utility Assessment (MAUM) method is used to weight the elements. This method is also explained in Chapter 3.

Part 2 - Judgmental Sources of Variance and Questionnaire File. The stage is now set for making estimates of each appropriate element or obtaining estimates from SMEs. The analyst designs analytic or judgmental instruments to gather needed detailed data. The basis for this design is shown in Table 1 in terms of judgmental sources of variance.

Judgmental sources of variance provide a useful way of conceiving the problem from an analytic standpoint. Empirical studies are concerned with varying independent variables to test their effects on dependent variables. However, analytic models must rely on the judgments of experts bolstered by available information inputs to formulate a TD/S concept and see it through specifications, contracting, development, deployment and fielding. It is only after major design decisions have been made that empirical testing can begin.

Conceiving of the problem in terms of judgmental variances can lead to useful ways of measuring, predicting or controlling sources of variance in the TD/S design and development process. This conception also leads to useful ways of formulating the instruments (questionnaires or interviews) required for their measurement, and of testing the reliability and validity of the analytic estimates. Illustrations of this approach are given in the review of DEFT and FORTE later in this chapter and the sample questionnaires in Appendix B, and in Chapter 3. The DEFT and FORTE experience demonstrates methods by which reliable and valid judgments may be economically obtained from SMEs.

Table 1 summarizes and gives examples of the two general variance sources, namely, variances associated with:

1. independent variables
2. dependent variables

The analyst selects the array of independent variables of interest that help form the TD/S concept, and the appropriate acquisition, transfer, safety and other dependent variables. He/she then tests alternative sets of independent variables for their relationships with the dependent variables. SMEs may be employed at this point to make the estimates or to cross-check the TD/S analyst's estimates. The acquisition, transfer of training, and job readiness estimates may require different SMEs than the accident probability estimates and the cost estimates.

Table 1

Judgmental Sources of Variance for the TECIT Analytic Component

Independent Variables	Dependent (Criterion) Variables
1. Training Program	1. Acquisition learning
2. Task Complexity	2. In-course Transfer of training
3. Physical and Functional Fidelity (Engineering Variables) 3.1 Motion 3.2 Visual 3.3 Auditory 3.4 Olfactory 3.5 Kinesthetic 3.6 Others	3. Safety 4. Job Readiness 5. Utilization ratio 6. Costs
4. Instructional Variables 4.1 Sequences 4.2 Cues 4.3 Feedback 4.4 Others	
5. Student Input 5.1 Knowledges 5.2 Skills 5.3 Attitudes	
6. Instructional Management 6.1 Instructor station training and utility 6.2 Instructor/trainee ratio 6.3 TD/S and WS scheduling 6.4 Downtime - based on reliability and maintainability	
7. Others	

The terminology used for the independent variables in Table 1 requires clarification. Rose and Wheaton (1985) in their development of the Device Effectiveness Forecasting Technique noted from their review of the literature the primacy of task difficulty as a dimension of transfer. The argument is simply that certain tasks are inherently more difficult to learn. For example, there may be more physical or mental "steps" in the learning process; they may require greater perceptual discrimination skills; or they may require greater psychomotor skills. A task profile is used to analyze the difficulty of the tasks.

The concepts of physical and functional fidelity are also adapted from Rose and Wheaton (1985). Physical fidelity is the extent to which the TD/S is perceived to be physically similar to the WS in its static state. Functional fidelity reflects the extent to which the TD/S reflects dynamic conditions similar to the WS in actual operation. Instructional variables are those that enhance learning.

Instructional management variances are expected to be related to the acceptability of a design and to utilization rates (Goldberg and Khattri, 1986). Experienced instructors may provide useful analyses of the utility of the instructor station and the feasibility of extended hours of training on the TD/S. Analysis of TD/S and WS scheduling may reveal bottlenecks or other constraints. Downtime due to TD/S and WS unreliability also need to be considered from a scheduling and implementation standpoint.

The following instruments are available or may be easily modified for use with each element of the effectiveness function:

1. DEFT Scales, with modification suitable for acquisition, safety, and transfer within the course. (See illustration in the next section of this Chapter and in Appendix B.)

2. FORTE scales, suitable for time to criterion within course transfer, but with modification also appropriate for acquisition learning and performance measures of transfer. (See illustrations later in this chapter and in Appendix B.)

3. Safety, job readiness and utilization ratio scales are presented in Chapter 3.

Combinations of scales for each effectiveness element are:

1. Acquisition - DEFT and FORTE

2. Safety - DEFT and TECIT scales

3. Transfer of training within the course - DEFT and TECIT scales

4. Job readiness - TECIT scale

5. Utilization ratio - TECIT scale

Joint consideration of all possible variances at one time is difficult. The challenge facing the TD/S development team is to define those variances that are most important to estimate for the intended application and to select or develop the instruments to measure those variances reliably. The TD/S development team may wish to develop priority listings of variances to be assessed in successive phases of TD/S development. The questionnaire file and the reviews of FORTE and DEFT illustrate how this is done.

From a research standpoint, there is also an analytic method variance that needs to be better explored to determine the conditions and applications for which an analytic method can provide reliable and valid estimates. This issue is discussed further in Chapter 4 Research Strategies.

REVIEW OF RELATED MODELS

A number of existing formal models concerned with TD/S development and forecasting contributed to our thinking in the development of TECIT. Parts of them have been adapted to TECIT and thus lend a background to important aspects of TD/S model development. These models are:

1. Device Effectiveness Forecasting Technique (DEFT Rose & Wheaton, 1984)

2. Forecasting Training Effectiveness (FORTE, Pfeiffer, Evans and Ford, 1985; Pfeiffer and Scott, 1985).

3. Comparison Based Prediction (CBP, Klein, 1985)

It should be noted that DEFT and CBP were reviewed in detail by Goldberg and Khattri (1986, Chapter 4) in a review of training effectiveness models. For this reason only essential features of these models are reviewed. FORTE was not reviewed in that report as the documents were not available to us at the time. FORTE is reviewed in greater detail in this report.

Device Effectiveness Forecasting Technique (DEFT)

DEFT emerged as a reconceptualization of the TRAINVICE models. These models were developed to predict TD/S

transfer to performance settings. The development of the TRAINVICE and DEFT models is reviewed in Goldberg & Khatri (1986) and TRAINVICE alone is reviewed in Knerr, Nadler and Dowell (1984) and Tufano and Evans (1982).

The DEFT authors (Rose and Wheaton, 1984) emphasize the importance of evaluating the training device within the framework of the training program in which it is embedded. The model is based on a program evaluation rationale or network of hypotheses which make explicit the dynamics of the cause-effect relationship. Figure 2 depicts that rationale. The model focuses on hypotheses that relate events at one stage of learning to those at the next stage of learning. A detailed program rationale of the deficit model is depicted in Figure 3 which relates the events of one stage to the next.

The analyst selects from three levels of analyses ranging from global to detailed: DEFT - I global; II - task level, and III - detailed subtask level. These levels of analysis are used at various phases of training device development, depending on the level of detail of the task analytic information available.

To use DEFT, the analyst enters responses to rating scales into a computer for each of the DEFT components. Four major analyses are conducted at each level:

1. Training Problem - (TP) is an estimate of the magnitude and difficulty in overcoming the performance deficit: the level and type of proficiency associated with the training objective and trainees' level of knowledge relative to this prior to using the device.

2. Acquisition Efficiency - (AE) takes into account the quality of training provided by the device and the extra device variables which affect acquisition of skills required to meet training objectives. Assessment is made of training principles and instructional features of the device.

3. Transfer Problem Analysis - (TRP) This is an estimate of the performance deficit that the trainees bring to the parent equipment after graduating from the training device. It assesses residual deficit and difficulty in overcoming this deficit. Also, physical and functional similarity between the device and equipment are assessed.

4. Transfer Efficiency Analysis - (TT) This is concerned with measuring the transfer of skills and knowledges learned from the device to the equipment. The analysis is an evaluation of the

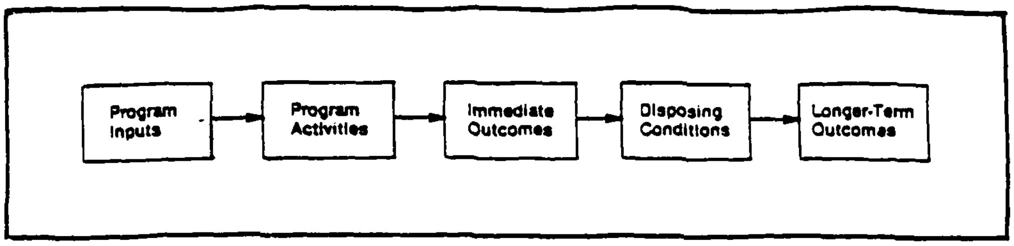

Figure 2. General model of the program rationale—DEFT
SOURCE: Rose & Wheaton, (1984)

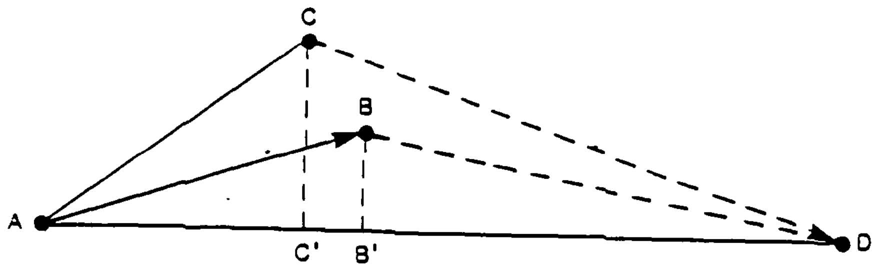

Figure 3 Deficit model of training device effectiveness. DEFT
SOURCE: Rose & Wheaton, (1984)

A = initial skills and knowledge of TRAINEE: performance on operational task prior to training on device (TD)

B = skills and knowledge of TRAINEE at completion of TD_1 regimen; criterion performance on TD_1

C = skills and knowledge of TRAINEE at completion of TD_2 regimen; criterion performance on TD_2

D = skills and knowledge needed to perform operational task; criterion performance on operational equipment

B', C' = skills and knowledge needed to perform operational task possessed by trainee after TD exposure; performance on operational equipment

AD = time, cost associated with learning D on operational equipment

AB, AC = time, cost associated with learning B, C on TDs

BD, CD = time, cost associated with learning D given learning on TDs

ABD, ACD = total time, cost associated with learning D for each TD

transfer principles that the device incorporates.

Table 2 shows the DEFT I indexes, formulae and range of values. The formulae differ slightly for DEFT II and III, averaging them over tasks. A copy of the DEFT I questionnaire in Appendix B shows the ratings for each scale. It can be noted that the DEFT component ratings are combined algorithmically into the various indexes.

The reliability and validity of DEFT have been explored in two studies. Rose and Martin (1984) conducted an initial assessment of DEFT. Six raters were used to determine the degree of inter-rater agreement. The raters evaluated three training devices: MK-60 gunnery trainer, burst-on-target trainer, and a maintenance procedures simulator. The authors claim that the data showed that DEFT I and III are internally consistent, but standard reliability indexes were not given. The FORTE review that follows provides additional data on the comparative reliability and validity of DEFT and FORTE.

In our opinion the DEFT model has made a substantial contribution to the literature of TD/S development and forecasting in organizing variables conceptually within a program evaluation rationale that defines and takes into account the training problem, performance deficit, learning difficulty, acquisition on the TD/S, task difficulty, physical fidelity, functional fidelity and transfer. We are incorporating these scales within TECIT. However, its use of rating scales instead of time and performance measures makes it difficult to validate in relation to empirical data and to interpret its ability to discriminate among TD/S design features. It makes no distinction between transfer, safety or job readiness and does not consider utilization in the effectiveness function. It is an empirical question as to whether or not the indexes employed (Table 2) for acquisition and transfer are universal. They imply a single function for combining various data elements rather than a family of functions specific to the individual case. Our point of view as to how to test these assumptions is to employ DEFT instruments as a structured line of questioning followed by questions specific to the individual case of acquisition or transfer. This approach is illustrated in Appendix B by a modified DEFT I questionnaire followed by a TECIT III set of questions for performance transfer for Simulated Combined Arms Training (SIMCAT) to a field exercise on the M1 Abrams Tank. When empirical transfer data become available, comparing the reliability and validity of the two methods will test the generality of the DEFT indexes.

Table 2

DEFT I Indexes

Training Problem (TP) = $\frac{\text{Performance deficit (PD) x learning difficulty (D)}}{100}$

Ranges from 0 to 100

Acquisition Efficiency (AE) = $\frac{\text{Rating}}{100}$

Ranges from .01 to 1.00

Acquisition (A) = $\frac{\text{Training Problem (TP)}}{\text{Acquisition efficiency (AE)}}$

Ranges from 0 to 10,000, with a low value indicating an "effective" device.

Transfer Problem (TRP) = $\frac{\text{RPD x RLD + AD}}{100}$

Where

RPD = Residual Performance Deficit

RLD = Residual Learning Difficulty

AD = Additional Deficits or Physical Similarity, Functional Similarity

Ranges from 0 to 200

Transfer Efficiency (TT) = $\frac{\text{Rating}}{100}$

Ranges from .01 to 1.00

Transfer (T) = $\frac{\text{TRP}}{\text{TT}}$

Ranges from 0 to 20,000, with a low value indicating an effective device.

Total Effectiveness (Σ) = A + T

Forecasting Training Effectiveness (FORTE)

The FORTE model was developed by Pfeiffer, Evans and Ford (1985) to simulate a variety of aviation training device evaluation outcomes by obtaining judgments from experienced instructors, supplemented by statistical modeling techniques. The model was specifically designed to explore sources of error variances threatening the sensitivity of device evaluations after a TD/S has been fielded. Variances explored in the two studies conducted so far include device features (i.e., visual and motion simulation), instructor leniency (i.e., easy, average, tough), task difficulty (i.e., easy, average, tough), and student ability (i.e., fast, average, slow). Input came from ratings made by flight instructor SME's on the FORTE rating scales (see Appendix B). These experts estimated trials-to-mastery in helicopters by trainees with and without prior simulator training.

The effects of these variables are estimated by two methods: interactive and additive. In the interactive method, the SME estimates the trials required for mastery for a number of training conditions. In the Pfeiffer et al. (1985) study, there were 27 conditions for the experimental group and 27 conditions for the control group. The training experts estimated trials for only eight conditions in each group. The rest were estimated by a regression subroutine in the model. Table 3 shows the eight conditions which were estimated by the SMEs.

The relative importance of the three variables (i.e., instructor leniency, task difficulty, student ability) is determined by the SMEs or the analyst. The parameters given in the model are shown in Table 4.

In the additive method, the averages of the trials-to-mastery for the experimental and control groups in the interactive method are used as a basis for estimating the deviations from the mean for each of the conditions. Six conditions were estimated for each group. The remainder were estimated by a computer model using the rules of conjoint measurement. These six conditions are shown in Table 5.

The model was validated using a concurrent validation design during an experimental evaluation of Device 2FG4C (SH-3) helicopter simulator in Jacksonville, Florida. Thirteen flight instructors currently involved in training the pilots took one-half hour each to complete both the additive and interactive rating methods. All four independent variables were utilized: device features, student ability, task difficulty, and instructor leniency. Trials-to-mastery was used as the dependent variable.

Results showed that the reliability for the 13 raters was $r = .97$ for the additive method and $r = .92$ for the

Table 3

Interactive Questionnaire Instrument for Estimating Trials-to-Mastery in the Forecasting Training Effectiveness Model (FORTE)

CONDITION	INSTRUCTOR	STUDENT	TASK	ESTIMATED TRIALS
1	Easy	Fast	Easy	____
2	Easy	Fast	Tough	____
3	Easy	Slow	Easy	____
4	Tough	Fast	Easy	____
5	Easy	Slow	Tough	____
6	Tough	Fast	Tough	____
7	Tough	Slow	Easy	____
8	Tough	Slow	Tough	____

SOURCE: Pfeiffer et al. (1985)

Table 4

Parameters for Weighting Trials-to-Mastery

Parameter	Relative Importance		
A	Instructors	Students	Tasks
B	Students	Instructors	Tasks
C	Tasks	Instructors	Students
D	Instructors	Tasks	Students
E	Students	Tasks	Instructors
F	Tasks	Students	Instructors

SOURCE: Pfeiffer et al. (1985)

interactive method. Inter-rater reliability using Pearson correlations to examine cross method variance was r = .92.

Validity analysis in Table 6 supports the accuracy of the modeled data for predicting the magnitude of the device feature effect. It is based on a comparison of FORTE data with empirical data of the field experiment.

The concurrent validity was estimated at r = .85 after each Pearson r was converted into a Fisher Z coefficient for averaging. These validity coefficients were for the two scales and the field experiment.

A linear extension of the model was developed by regression analysis of the simulated data. This analysis, shown in Table 7, indicated that the smallest amount of variance is attributable to the device features (.07). The other three variables combine to yield .90 of the variance.

That task difficulty accounted for the largest part of the variance (.42) is consistent with its importance in the DEFT concept. Instructor leniency (.21), a measure of criterion unreliability, suggests the need for more consistent measurement of performance.

A second study by Pfeiffer and Scott (1985) examined the separate and joint effects of visual and motion simulation on pilot flight performance of the SH-3 helicopter flight simulator. Both experimental and analytic methods were employed. The analytic methods used were DEFT I and II and FORTE enabling a comparison to be made of the two methods. (See Appendix B for the questionnaires.) This report was interested in determining the accuracy with which it is possible to predict transfer using the DEFT and FORTE analytic models.

SMEs were two instructors from the Naval Training Systems Center. Rater 1 was familiar with DEFT, FORTE and the device. Rater 2 was unfamiliar with DEFT and FORTE but very familiar with the device.

Pfeifer and Scott (1985) evaluated four device features: visual only (VISNLY), visual and motion (VISMOT), motion only (MOTNLY) and no motion - no visual (NVSMOT). Results in Table 8 show that the inter-rater reliability for DEFT II was much higher (.81 to .97) than for DEFT I (.39 to .72). DEFT II acquisition scales had somewhat higher reliability (.96 and .97) than DEFT II transfer measures (.81 to .96).

Table 9 shows that the additive method, FORTE II, showed higher reliability than the interactive method, FORTE I. The reliability for FORTE I was in the .70s and FORTE II in the .90s.

Table 10 shows the modeled and actual transfer ratios by device feature. FORTE was much more accurate than DEFT in

Table 5

Additive Questionnaire Instrument for Estimating Trials-to-Mastery in the FORTE model

IF AN AVERAGE STUDENT REQUIRES *N* TRIALS TO LEARN TO MASTERY, HOW MANY TRIALS WILL A ... FAST LEARNER REQUIRE?
... SLOW LEARNER REQUIRE?

IF AN AVERAGE INSTRUCTOR REQUIRES *N* TRIALS TO TRAIN STUDENTS, HOW MANY TRIALS WILL ... AN EASY INSTRUCTOR NEED?
... A TOUGH INSTRUCTOR NEED?

IF *N* TRIALS ARE NEEDED FOR AVERAGE TASKS, HOW MANY TRIALS WOULD... ... AN EASY TASK REQUIRE?
... A TOUGH TASK REQUIRE?

Note - *N* is based on mean trials from the interactive method rounded to the nearest whole number.

SOURCE: Pfeiffer et al. (1985)

Table 6

Modeled and Actual Trials-to-Mastery in the SH-3 for Two Conditions of Prior Training in Device 2F64C

TYPE ESTIMATION	VISUAL MOTION	MOTION ONLY	DIFFERENCE
INTERACTIVE METHOD	4.54	5.64	1.10
ADDITIVE METHOD	4.69	5.67	0.98
FIELD EXPERIMENT	4.68	5.41	0.73

SOURCE: Pfeiffer et al. (1985)

Table 7

Relative Contribution of Independent Variables
to Estimate Trials Needed for Mastery in Aircraft
(Values are Based on Simulated Data)

INDEPENDENT VARIABLE	CORRELATION r	VARIANCE r^2
Device Feature	.26	.07
Instructor Leniency	.46	.21
Student Ability	.52	.27
Task Difficulty	.65	.42

SOURCE: Pfeiffer, Evans & Ford (1985)

Table 8

Reliability of DEFT Scales for

The Average of Two Raters Using Tasks

From "A" Stage Training

SCALE	N ITEMS	RELIABILITY
DEFT I		
VISMOT	8	.72
VISNLY	8	.55
MOTNLY	8	.39
NVSMOT	8	.60
DEFT II ACQUISITION		
Performance Deficit	16	.97
Learning Difficulty	16	.97
Quality of Training Acquisition	16	.96
DEFT II TRANSFER		
Residual Learning Difficulty	16	.96
Physical Similarity	16	.85
Functional Similarity	16	.81
Quality of Training Transfer	12	.92

SOURCE: Pfeiffer & Scott (1985)

Table 9

Reliability of FORTE Scales for the Average of Two Raters

SCALE	N ITEMS	RELIABILITY
<u>FORTE I</u>		
VISMOT	8	.73
VISNLY	8	.73
MOTNLY	8	.74
NVSMOT	8	.69
FLYNLY	8	.77
<u>FORTE II</u>		
Student Ability	10	.98
Instructor Leniency	10	.99
Task Difficulty	10	.99

SOURCE: Pfeiffer & Scott (1985)

predicting the transfer ratios. FORTE II (the additive method) proved to be most accurate for forecasting the effectiveness of the various device features. The transfer ratio employed was the proportion of trials saved on the helicopter.

Table 11 shows that the convergent validity combining DEFT and FORTE transfer coefficients averages r = .92. Concurrent validity for DEFT transfer is r = .55, and for FORTE is r = .78. Apparently both methods contributed independently to predicting transfer.

It should be noted in Table 10 that the actual transfer ratio for the no-visual/no-motion group was higher than for the motion only group. This finding was not predicted by DEFT I, II or FORTE I. The authors suggest that the DEFT model does not properly combine physical and functional fidelity scales to yield an appropriate transfer coefficient. They also suggest that DEFT scaling should be modified to include such scales as trials-to-mastery, time-to-mastery, the transfer ratio or the transfer effectiveness ratio.

In our opinion, FORTE has made major contributions to the TD/S forecasting literature in devising the concept of judgmental sources of variance, methods for measuring them, coupling them with statistical estimating routines, and demonstrating the reliability and validity of the methods for forecasting empirical data. In contrast to the DEFT rating scales and formulae, their scales of measurement (time and trials to criterion) readily lend themselves to analysis in relation to empirical experiments. The applications of FORTE so far have been limited to aiding in device evaluation designs by estimating sample sizes needed for various levels of statistical significance and power; estimating the magnitude of variance sources, estimating the masking effects extraneous variance sources (i.e., task difficulty, student variance, instructor leniency) may have on TD/S characteristics (i.e., visual and motion simulation) and demonstrating the ability of their measures to discriminate among TD/S characteristics. Unlike DEFT, however, FORTE has not addressed acquisition learning on the TD/S or TD/S design and has not used a structured line of questioning to channel the SMEs thinking about the training program, physical and function fidelity issues and other matters. The FORTE authors (Richard Evans, Personal Communication, April 1986) and the authors of this model believe there is much in DEFT and FORTE worth considering in further research on analytic methods for TD/S.

Comparison-Based Prediction (CBP)

Klein Associates' (1985) Comparison-Based Prediction (CBP) is an approach intended to be applied to TD/S early in the design sequence. This method does not require

Table 10

Comparison of Modeled and Actual

Transfer by Device Feature

Using Tasks from "A" Stage Flight Training

DEVICE FEATURE	MODELED TRANSFER COEFFICIENT				ACTUAL TRANSFER RATIO
	(DEFT I)	(DEFT II)	(FORTE I)	(FORTE II)	(TR)
VISMOT	.92	.84	.37	.34	.29
VISNLY	.88	.82	.33	.31	.27
MOTNLY	.82	.80	.26	.16	.20
NVSMOT	.79	.77	.24	.20	.25

SOURCE: Pfeiffer & Scott (1985)

Table 11

Validity of DEFT and FORTE

for Estimating Transfer of Training

MODEL	TYPE VALIDITY	RANGE	MEAN
DEFT AND FORTE	Convergent	.81 - .99	.92
DEFT	Concurrent	.45 - .63	.55
FORTE	Concurrent	.68 - .87	.78

SOURCE: Pfeiffer & Scott (1985)

operational data from the system under design; it may operate with information from sources similar to the TD/S. CBP utilizes structured expert opinion. CBP is "...a method of reasoning by analogy, where an inference is made for one object or event based upon a similar object or event..." (Klein 1985, pp. 1-4).

The methodology is described as follows:

<u>Elements of the CBP methodology</u>

1. Target Case A
2. Target Variable: T
3. Target Value: T(A)
4. Subject Matter Expert (SME)
5. Comparison Case(s): B
6. Causal Factors (from which high drivers are selected)
7. Scenario
8. Strategy
9. Comparison value: T(B)
10. Audit Trail

<u>Steps in using CBP</u>

<u>Phase I: Set up the problem:</u>

1. Specify the device (a) for which cost effectiveness is being predicted.
2. Define the measure (T) of that cost or effectiveness. This is the variable to be predicted.
3. Identify the major causal factors (high drivers) that affect T(A).
4. Define the context for the prediction. This includes when and where and how the device will be used.

<u>Phase II: Select Specific Resources</u>

5. Identify comparison devices.

6. Examine the CBP strategies to select the most relevant one.

7. Choose knowledgeable subject matter experts.

Phase III: Collect the Data

8. Determine, with the SME, the comparison value T(B).

9. Examine the difference between A and B, and estimate the effect of these differences on T(B).

10. Adjust the value of T(B) to account for the differences between A and B.

Phase IV: Make the Prediction

11. Determine the value for T(A) from this adjustment.

12. Document the process to leave an audit trail. This aids in evaluating this decision or in revision as further development takes place.

The steps outlined above for using CBP are not to be taken as rigidly sequential. Alternative strategies can be used depending upon time constraints, the number of comparison cases, availability of data, and identification of SMEs. The alternative strategies include:

1. Global strategy - One SME is interviewed and presented with all relevant data on A, including a list of high drivers. The SME makes a prediction for T(A) based on his/her knowledge of T(B).

2. High driver strategy - The SME details how A and B differ from one another. With a checklist of high drivers, the SME compares the two devices on these high drivers and how much difference they effect. The sum of these estimates is then calculated.

3. Multiple comparison strategy - Several comparison cases are initially used, then the choice is narrowed down to two or three.

4. Convergence strategy - Use of multiple comparison strategy as well as use of SMEs multiple strategy. When using multiple comparisons, the

SMEs should be asked to rate only the device with which they are familiar. If they are experienced with more than one, the list of causal factors should be reduced to make it less confusing.

5. Cumulative strategy: The SMEs can be added and interviewed one-by-one until enough agreement is achieved.

The authors give further guidance on the collection and analysis of data and on documenting the process.

According to Klein (1985) CBP has a number of characteristics which make it useful to apply in the early stages of training device development. It does not require extensive data from the device about which predictions are to be made; predictions are derived from operational experience; it uses structured expert judgment; it asks for judgements relative to similar cases; and it leaves an audit trail of the prediction process.

According to the authors, CBP has been developmentally tested in predicting such measures as time saved in training and effectiveness of training. CBP has been applied to automotive maintenance trainers, VideoDisc gunnery simulators for tanks (VIGS), and trainers for self propelled howitzer operations and maintenance (HIP). The author indicates that CBP methodology has been compared with actual test results of effectiveness of training devices at George Mason University. The results, as yet not published, yielded a correlation of .90 between CBP predictions and test results. Another study noted that training personnel showed greater confidence in predictions using the CBP methodology as compared with their own unstructured judgments. (Klein, 1985)

In our opinion, CBP has identified an area worth considering and formalized a process for doing so: the situation in which there is a similar TD/S from which estimates can be made for a newly developing TD/S. However, a great deal of the process as described represents defining the problem, much like any other problem solving process. DEFT and FORTE have not given explicit attention to this part of the process. In TECIT, we have formalized the process as the Problem Definition and Analysis Component (see Chapter 2) and incorporated consideration of similar and predecessor TD/S. Presumably, DEFT and FORTE could use similar TD/S as part of the input to SMEs. CBP's greatest shortcoming appears to be in its measurement approach. The author lists many variables which can be addressed, but does not organize the variables conceptually as does DEFT or FORTE (i.e., TD/S acquisition, transfer, physical fidelity, etc.). In the studies reviewed, only a few variables are addressed in each study, giving a very limited picture of

predicted effectiveness. The method does not easily lend itself to statistical estimates of reliability and validity. However, coupling it with methods such as those used by FORTE could overcome these problems.

Comparison of TECIT AND Other Models

Table 12 compares TECIT with DEFT, FORTE and CBP. This table summarizes all relevant comparative aspects of these models.

ADDITIONAL DEVELOPMENTS NEEDED FOR TECIT

TECIT is a generic, modular, multi-purpose model adaptable to a variety of entry points and applications. Further development of the model is needed in a number of areas. These areas are listed in the following order of priority:

1. Users Application Guide. Various elements of the model are more appropriate to one type of application than another. For example, conceptual design applications would rely more on baseline sources, selecting and monitoring contract developments, while fielding applications would rely more on forecasting methods relevant to installation, piloting, and empirical validation. TD/S developed for safety reasons and criterion referenced (work sample) TD/S require somewhat different consideration. The development of valid and reliable performance measures and methods for combining them may be an important early consideration in TD/S design and in the WS exercise in designing and restructuring training. Technology transfer, exportability, multi-course applications, career sequences, formal school vs. OJT sequences, and system vs. non-system applications require illustration and guidance.

   Finally, applications to training of different types of personnel (such as tank commanders, gunners, driver, pilots, navigators, maintenance, or supply) may require differing emphasis on the configuration of acquisition, transfer, safety, job readiness and instructional management and the independent variables. Flow charts, illustrations and guidance would be helpful to users.

   An expanded questionnaire file for assessing sources of analytic variance related to various applications would be a useful aid.

2. Research Guide. A general research strategy is outlined in this report in brief. A research guide

would expand on this strategy showing how various analytic and empirical study designs can be formulated and carried out. The following should be considered: coupling applications and research; research on model iterations; reliability and validity of analytic estimates as a function of application information input (WS development and training program development), analyst characteristics and SME characteristics; cross-sectional vs. longitudinal designs; reliability methods; concurrent, convergent discriminant and predictive validity designs; designs coupling empirical and analytic methods; analytic designs relating the independent variables (such as student characteristics, physical and functional fidelity and instructional management) to dependent variables; exploration of the relationships of additional judgmental variance sources such as instructor leniency, instructor quality, objective vs. subjective performance measures, sources of criterion unreliability, student experience, student quality, team variance, nested task difficulty studies; approaches for validating safety, job readiness and instructional management scales; the multivariate structure of independent and dependent analytic and empirical variance sources; the analysis of acquisition and forgetting functions in relation to relearning, knowledge and skill integration, and the planning and implementation of refresher training, skill retention training and cross-training in career sequences; the development and incorporation of useful analytic and empirical databases and meta-analyses; available computer routines and their uses; hypotheses generation vs. hypothesis testing approaches and other topics as appropriate.

3. Computerization. It should be apparent that even though manual applications are feasible, applications and research would benefit from computerization of TECIT. A software structure of problem definition and analysis, applications, graphic aids, questionnaire generation/revision and statistical and calculation routines would aid analysts in formulating their approach. Analyst and SME data entries would be made directly on the computer. Data storage would provide the audit trial necessary for model iterations and research. Computerization is listed as a third priority as it would be most useful to formulate the details of the users guide and research guide before developing the supporting computer software. Furthermore, a common computer system is being developed for certain applications for the Army (Personal Communication, D. Haggard, March, 1986) and it may be appropriate to wait until this new system is available.

As applications and research evolve, the extent to which the model has generic qualities can be assessed and adaptations and revisions made.

Table 12

Comparison of TECIT and Other Models

	TECIT	DEFT	FORTE	CBP
1. Conceptual orientation	Multi-purpose method. Defines modeling applications at various phases of the TD/S, WS, and training program life cycles. Uses measures of acquisition learning, transfer, instructional management, safety and job (battle) readiness for forecasting and to complement empirical studies. Like FORTE, assumes a family of application specific transfer functions.	Forecasting acquisition learning and transfer of training based on a program evaluation rationale and transfer theory. Appears to be intended for use primarily in the TD/S design phase. Organization and scoring of acquisition and transfer indexes implies a unitary functional relationship rather than a family of relationships.	An aid in designing field evaluations of flight simulators. It has been used only in the fielding phase of a TD/S. Concerned with transfer of training. Assumes a family of transfer functions specific to particular applications.	Uses similar TD/S in the design phase when no empirical data are available. SMEs predict based on comparison from similar to proposed TD/S. An aid in problem definition but not concerned with data organization.
2. Applications in TD/S life cycle phases				
2.1 Design phase applications	Yes. Is a TD/S needed? What kinds? Concept formation, contract guidance	Yes, but unarticulated in the model	No, but possible	Yes, primary purpose
2.2 Fielding phase applications	Yes, aid in field evaluation and judgmental measures of effectiveness	Unknown	Yes, primary purpose is aiding design of field evaluations	No
3. Joint use of analytic and empirical data	Yes, proposed application	No	No, but possible	No

Table 12 (con't)

	TECIT	DEFT	FORTE	CBP
4. Problem definition analysis and information gathering process	Yes, a specific systematic component	Not explicit	Not explicit	Yes, a major part of the process
4.1 Definition of training spectrum & expected range of applications	Yes, in #4	No	No	No
4.2 Includes database, predecessor or similar TD/S	Yes, all in #4	No, but possible	No, but possible	Similar TD/S only. Others possible.
5. Organization of data				
5.1 Considers acquisition learning on TD/S	Yes, important	Yes, important part of concept and indexes	No, but possible	No planned data organization. Obtains 1 or 2 data points depending on specific problem definition.
5.2 Considers transfer of training	Yes, important	Yes, important part of concept and indexes	Yes, primary purpose	
5.3 Measures used in acquisition and transfer	Time to criterion, total training time, and performance for acquisition and transfer	Rating scales	Time or trials to criterion in transfer, but adaptable to performance measurement	

Table 12 (con't)

	TECIT	DEFT	FORTE	CRP
5.4 Safety, instructional management, job readiness	Yes	No	No	No
6. Reliability and validity methods	Yes. Propose to adopt and extend FORTE approach	Yes, reliability based on rating scales. Rating scales pose a problem in relating DEFT data to empirical data for validation studies. See FORTE vs. DEFT comparative study review.	Yes, employs SME judgmental variances & statistical routines to obtain inter- and intra-rater reliability, variance estimates, discrimation, accuracy, concurrent, convergent and predictive validity. Two studies show promising results. Criteria for validity are empirical transfer studies.	Yes, but reports of methods not found. However methods may be limited because of interview method and limited data obtained
7. Designed for joint application & research	Yes	No	Yes	No
8. Research Strategy & Validation Plan	Yes, based on applications, computerized audit trail, and special research projects	Yes, but not necessarily related to applications	Under discussion. Not yet documented	None found

Table 12 (con't)

	TECIT	DEFT	FORTE	CBP
9. Audit trail	Yes, for applications & research. Detailed with regard to problem definition, SMEs, study design, analytic method and findings.	Unknown	Yes, for applications and research. Detailed with regard to SMEs, study design, analytic method and findings.	Yes, for applications. Detailed with regard to SMEs method & findings.
10. Computerized	Proposed as one of next steps in development. Manual and computer methods proposed for problem definition analytic methods and statistical analysis, extending the FORTE approach.	Yes. Questions, scoring indexes and data summaries all on computer. Manual application possible but difficult.	Yes. Used for questionnaires, design, presentation to SMEs, and statistical analyses. Manual applications also possible.	No. Not proposed as data organization & volume does not require a computer; flexible interview method does not lend itself to computerization.
11. Cost Analysis	Yes. Life cycle cost model and cost-effectiveness decision methods	No	No	No cost model, but costs may be one of data items gathered in comparing similar and proposed TD/S

Chapter 2

TRAINING EFFECTIVENESS OF TECIT: PROBLEM DEFINITION

INTRODUCTION

The training effectiveness of TECIT has two major components:

Component 1: Problem definition: the training spectrum, context, purpose, information gathering and baseline analysis.

Component 2: Analytic forecasting and judgmental methods.

This chapter presents in detail Component 1 of the training effectiveness submodel of TECIT. Each section of the chapter explains the rationale, presents the forms to be used, the applications that can be made, and research uses of the information. Chapter 3 discusses Component 2 in detail.

TRAINING SPECTRUM ANALYSIS (Form 1)

The training spectrum refers to the range of applications anticipated for the TD/S. For example, the TD/S may be developed for system training or for non-system training; for one course or several courses; for one or a number of sites; and for use by a variety of personnel. Documenting the range of intended applications aids in selecting candidate forecasts. Form 1 is used for this analysis.

Different uses of the TD/S will have a differential impact on the design of the forecasting study. For example, if the TD/S is to be used for one WS and the same course in a number of different sites, then the differences in the student body and instructional program will have to be taken into account. However, if the TD/S is to be used for different courses, personnel and WS, then different sets of criterion measures will have to be established for the analysis. Thus, the training spectrum has to be documented before forecasting criterion metrics can be established. One analysis should be made for each major application anticipated.

For research purposes, Form 1 will provide a great deal of contextual information and the rationale behind the forecasting analyses made.

FORM 1: TRAINING SPECTRUM ANALYSIS

This form is to be completed by the TD/S analyst.

The Training Spectrum is the range of applications anticipated for the TD/S and helps guide the forecasting analysis.

First answer questions 1-8 below. Next append a detailed analysis to answer the sample questions listed in 9 and 10.

1. Analysts Name ____________________ Date Completed ______________

2. Training Device or Simulator (TD/S) name & number: ________________

____

3. Brief description of the TD/S. Attach or reference detailed functional description if available.

____

____

4. TD/S is to be used for (check one) ___System Training, ___ Non-System Training.

5. School name(s) and locations; job site location(s)

____

____

6. Course(s) title(s) and number(s)

____

____

7. MOS s of personnel to be trained

	a. Operators	b. Maintenance	c. Other
7.1 Regular Army	-------	-------	-----
7.2 Reserves	-------	-------	-----

8. Weapon System(s)____

____

9. Append a detailed analysis to answer questions such as the following: (a) Where is the TD/S expected to be placed within each formal course for which it will be used? For each course, what prerequisite training will be required. Give the type of prerequisites and hours of instruction. (b) Are there other TD/S's available or in development that may impact entry level skills of trainees or teach some of the same or other tasks in the sequence? (c) When and for how long will the WS by used for training? Before or after the TD/S? (d) What is the training-to-job-to-training sequence in this career such as initial training, OJT, refresher or transition training? In which parts of the career sequence will the TD/S be used?

10. From this analysis of the Training Spectrum and priority applications, list candidate forecasting analyses to be made.

CONTEXT - LIFE CYCLE DEVELOPMENT PHASES OF THE WEAPON SYSTEM (WS) AND TRAINING PROGRAM (TP) (Form 2)

The phases of development of WS and TP for system and non-system training give further guidance for the analysis. For example, when a WS is in the Conceptual or Demonstration and Validation Phase and the TP is in the Analysis or Design Phase, there are no data available about them and greater uncertainty about the impact they may have on the TD/S design. On the other hand, there may be greater flexibility in weighing the relative merits of matters such as the following:

1. Is a TD/S needed? One or a family of TD/Ss? Would a family of TD/Ss obviate scheduling problems? What type(s) of TD/S(s) should be developed?

2. Where and how should certain enabling skills be taught? In the classroom setting? With or without media support? On the TD/S?

When the WS has reached full scale development there is less risk that there will be changes in the WS that could call for changes in the TD/S. When the TP has reached the design phase, preliminary WS time estimates and performance measures may be available that will aid in forecasting transfer of training. When the WS has been fielded and the TP implemented, WS time and performance criteria are available for use in forecasting transfer of training. The risk is much lower (but not zero) that there will be significant changes in the WS or the TP.

In non-system training, some WS and TP may be fielded while others are in earlier phases of development. The analyst may then give primary attention to fielded WS and TP because of the data available for forecasting transfer of training. Furthermore, the design of the TD/S may result in adopting characteristics useful for the most demanding WS and TP application.

LIFE CYCLE PHASE OF THE TD/S AND PURPOSES OF THE ANALYSIS (Form 3)

The analyst checks the development phase and major purposes of the analysis on Form 3 and comments as appropriate. Note that the purposes of the analysis correspond to the development phases of the TD/S, reflecting the distinction between the early phases when major conceptual, design and cost decisions are being made and later phases, when the "metal is bent" - that is, when forecasting, "fine-tuning" the design, and planning utilization and empirical studies are paramount concerns.

FORM 2: LIFE CYCLE DEVELOPMENT PHASES OF THE WEAPON SYSTEM(s) (WS) AND TRAINING PROGRAM(s) (TP) FOR WHICH THE TD/S IS BEING DEVELOPED

The Life Cycle Phases of the WS and TP for which the TD/S is being developed establish the purposes of the analysis (design vs. forecasting). When the WS and TP are in advanced phases of development or fielding, data from them may be used to aid in TD/S design and forecasting.

1. Analysts' name ____________________ Date completed ________________

2. TD/S name and number ____

3. TD/S developed for (check one) ___ System training; ___ Non-system training.

4. Enter WS (use one for system training, all for non-system training) for which the TD/S is to be used. Add additional pages if needed.

   4.1 _________ 4.2 _________ 4.3 _________ 4.4 _________

5. Check the one corresponding life cycle development phase for each WS.

5.1 Conceptual	______	______	______	______
5.2 Demonstration and Validation	______	______	______	______
5.3 Full Scale Development	______	______	______	______
5.4 Production and Deployment	______	______	______	______
5.5 Fielded	______	______	______	______

6. Check the one corresponding life cycle development phase of the training program for each WS. If more than one TP, attach a separate copy for each.

6.1 Analysis	______	______	______	______
6.2 Design	______	______	______	______
6.3 Develop	______	______	______	______
6.4 Implement	______	______	______	______

7. Comments: ____

____

____

____

FORM 3: LIFE CYCLE PHASE OF THE TD/S AND PURPOSES OF THE ANALYSIS

The Life Cycle Phase of the TD/S relates to the major purposes of the TECIP analysis.

1. Analyst's name ____________ Date completed ______________________

2. TD/S name and number ____

3. Life cycle development phase of the TD/S (check one)

   ___3.1 Conceptual

   ___3.2 Demonstration and Validation

   ___3.3 Full Scale Development

   ___3.4 Production and Deployment

   ___3.5 Fielded

4. Major purposes of the Analysis: (check all that apply)

DESIGN (Primarily Conceptual and Demonstration/Validation Phases)

___(1) concept analysis and development - should a TD/S be developed? If yes, what types?

___(2) evaluating alternative design proposals and selecting among them.

___(3) working with contract developers to optimize design effectiveness and costs

___(4) acceptance testing

FORECASTING (Primarily Full Scale Development, Production/Deployment and Fielded Phases)

___(5) forecasting acquisition learning

___(6) forecasting transfer of training effectiveness

___(7) forecasting and planning training deployment and time

VALIDATION

___(8) designing empirical studies of acquisition learning and transfer of training

___(9) validation of the model - relating forecasts to empirical data

___(10) Other (explain in comments)

5. Comments ____

____

____

INFORMATION GATHERING (Form 4)

Once candidate analyses and the context and purposes of the TECIT analysis are identified, information is obtained about the following:

1. WS(s)
2. the training program(s)
3. the TD/S
4. predecessor TD/S
5. similar TD/S

Form 4 shows the format for organizing the information. This form is useful in a number of ways. First, it alerts the analyst to the various types of information to seek out and assemble depending on the purposes of the analysis and the phase of development of the WS, TP and TD/S. For example, if the purpose of the analysis is to develop TD/S design concepts in relation to a WS in the conceptual or development phase and a TP in the analysis and design phase, the analyst should seek out information on the threat scenario, WS concept functional description, drawings or mock-ups (item 3.1 on Form 4); and the TP task/subtask/skill analysis, design concept, and performance objectives (items 4.1, 4.2 and 4.4 on Form 4). The availability of a predecessor TD/S or similar TD/S (items 6 and 7 on Form 4) can also be determined as possible aids in developing the TD/S design concept (items 5.1 through 5.5 on Form 4). If the WS and TP are in advanced phases of development or fielded, the analyst should also be able to use 3.2, 3.3 and 4.3 through 4.6 (Form 4) and be able to more clearly relate WS and TP time and performance measures to TD/S time and performance measures.

Second, sources of data useful for preliminary estimates of transfer or acquisition measures may be found. When time or performance measures are available for a fielded WS and TP (4.4 and 4.6) these data may be used to obtain preliminary estimates of acquisition or transfer. Predecessor and similar TD/S and the database may also be helpful in this regard.

Third, Form 4 aids in identifying information and observations that will be needed by SMEs to make forecasts using the TECIT analytic model. In general, the type and amount of information should be expected to differ in terms of the familiarity of the SMEs with the WS, TP and TD/S concept and the availability of information at various phases. For example, in the early conceptual phases for TD/S, the WS and TP, the analyst may select SMEs with high levels of expertise in engineering design of WS and TD/S, human-factors, training and TD/S learning designs, and expert instructors. This mix of SMEs may continue throughout the development phases of the WS, TP, and TD/S. In contrast, after the WS, TP, and TD/S have been developed,

FORM 4: INFORMATION GATHERING

Information gathering is carried out to assess information needs and for presentation to SMEs. The information need not be gathered all at one time.

Directions: Check all that apply and describe specific information as indicated.

1. Analyst's Name ________________________ Date completed ___________

2. TD/S name and number ____

3. Weapon System(s) name and number ________________________________

Attach extra pages if more than one. Complete for candidate analyses only.

___ 3.1 Threat scenario, concept, functional description, drawings, mock-ups

___ 3.2 Observation of prototype weapon system

___ 3.3 Observation of operational, fielded weapon system

___ 3.4 Not needed. All SME s are familiar with the WS

___ 3.5 Other - specify below

Describe the specific information about the WS to be presented to SME's to aid them in making forecasts.

____

____

____

____

4. Training Program(s) name(s) and number(s) ___________________________

Attach extra pages if more than one. Complete for candidate analyses only.

___ 4.1 Task/subtask/skill analysis

___ 4.2 Design concept

___ 4.3 Description of developed training program for piloting

___ 4.4 Performance objectives and/or measures

___ 4.5 Description of fielded training program

___ 4.6 Description of how, when and how long the WS will be used in training and on the job. NOTE: This is a preliminary source of WS time and may be used in certain transfer formulae.

FORM 4: INFORMATION GATHERING (con't)

___ 4.7 Not needed. All SME s are familiar with the training program(s)

___ 4.8 Other - specify below

Describe the specific information about the training program(s) to be provided to SME s to aid them in making forecasts or to consider design questions such as where enabling objectives are to be taught, scheduling, a family of TD/S, etc.

____

____

____

____

____

5. Training Device/Simulator (TD/S)

___ 5.1 TD/S concept, functional description, drawings, mockups

___ 5.2 Task/subtask/skill or exercise analyses relevant to the TD/S as opposed to the training program in general for each TP

___ 5.3 Courseware appropriate to each TP

___ 5.4 TD/S performance objectives and/or measures

___ 5.5 Descriptions and analysis of instructors stations, response recording, instructor roles in such matters as selecting exercises, providing feedback, rating performance

___ 5.6 Observation of a prototype TD/S

___ 5.7 Observation of a fielded TD/S

___ 5.8 Brief "walk through" a TD/S excercise

___ 5.9 Description of how, when and how long the TD/S will be used in training and on the job. NOTE: This may be a preliminary source of TD/S time and may be used in certain transfer formulae

___ 5.10 Other - specify below

Describe the specific information about the TD/S to be presented to SME s to aid them in making forecasts.

____

____

____

____

6. Predecessor TD/S - A predecessor TD/S is one used in training with a predecessor WS or one that is being improved.

6.1 Is there a predecessor training device or simulator?

___ 6.1.1 No

___ 6.1.2 Yes, give name & number___________________________

6.2 Check the information available about the predecessor TD/S

___ 6.2.1 Drawings, description, photos, films, mock-ups

___ 6.2.2 Observation and "walk-through" of TD/S

___ 6.2.3 Description of time and performance measures

___ 6.2.4 Description of how, when and how long the predecessor TD/S was used in training and on the job

___ 6.2.5 Information about the instructor station and its utility

___ 6.2.6 Transfer of training data

___ 6.2.7 Other - specify below

6.3 Is it included in the Data Base? (see Appendix)

___ 6.3.1 Yes (Note: Data is useful for comparison to other TD/S).

___ 6.3.2 No

6.4 Usefulness of prior data (6.2.3, 6.2.4 and 6.2.6) depends on similarity of tasks, skills and excercises from predecessor to new TD/S, and changes in the threat scenario, performance objectives, measures, and new technological developments. Analyze the similarity of tasks/skills/excercises in the new TD/S vs. the predecessor TD/S. See Form 5 as an example to guide the analysis.

Describe the specific information about the predecessor TD/S to be presented to SME s to aid them in making forecasts.

____

____

____

____

____

7. Similar (not Predecessor) TD/S. Note: Information useful for CBP method.

7.1 Is there a similar TD/S? Check the Data Base (see Appendix) and other sources for candidates.

___ 7.1.1 No - (Stop)

___ 7.2.1 Yes - name(s) and number(s)______________________

If more than one, complete an additional form for each one.

7.2 Is it included in the Data Base?

___ 7.2.1 Yes - Note: Data useful for comparison

___ 7.2.2 No

7.3 In what ways are they similar?

___ 7.3.1 Both are primarily concerned with safety and procedural training

___ 7.3.2 Both simulate battle conditions that might otherwise be infrequently encountered

___ 7.3.3 Both give experience in maintenance tasks that might otherwise not be possible within limited training time or limited job experience.

___ 7.3.4 Both are designed for gunnery training

___ 7.3.5 Both use similar time or performance measures

___ 7.3.6 Most of the tasks and skills appear similar

___ 7.3.7 Other - specify ______________________________________

____

____

7.4 Indicate ways in which they are dissimilar.

____

____

____

____

____

____

7.5 Check sources and determine the information available about the similar TD/S(s).

___ 7.5.1 Description, drawings, photos, films, mock-ups, etc

___ 7.5.2 Observations and "walk-through" of TD/S

___ 7.5.3 Description of time and performance measures

___ 7.5.4 Description of how, when and how long the similar TD/S was used in training or on the job

___ 7.5.5 Transfer of training data. Note: Useful for comparison

___ 7.5.6 All information in hands of expert sources

___ 7.5.7 Other - specify

Describe the specific information to be provided to SME's to aid them in making forecasts.

____

____

____

____

____

____

more attention may be given to SMEs with expertise in training design and to "expert" instructors. Furthermore, different SMEs may be called for if there is a predecessor or similar training program. Different SMEs may be selected at various phases. See Form 7.

For research purposes, Form 4 documents the information available for analysis and for presentation to SMEs.

TASK/SUBTASK/SKILL COMPARISON (Form 5)

The Task Analysis Comparison Chart was devised to compare predecessor and proposed TD/S, but may also be adapted to task comparisons on the TP vs. TD/S and, where sufficient information is available, to a similar TD/S.

The tasks, subtasks, skills and exercises must be available in sufficient detail to make comparisons. The degree of similarity/dissimilarity can give leads to the credence to be given to baseline forecasting methods as opposed to analytic forecasting methods. Scoring of Form 5 is envisioned on a task by task basis, for all tasks combined and for all applications (i.e., courses, sites, WS) separately and combined. Computer programs would be useful for more detailed and complex analyses to aid the analyst in compiling and analyzing the data.

For research purposes, Form 5 provides further documentation of information gathering and analysis and the rationale for the relative weight to be given to baseline analysis depending on the degree of similarity of predecessor and new TD/S. It also summarizes information used by SMEs in making analytic forecasts.

BASELINE DATA SUMMARY (Form 6)

When the WS and training program are in advanced stages of development or have been fielded, or when there is a predecessor TD/S, a similar TD/S or a suitable data base, data elements may be available relevant to acquisition learning and transfer of training. The analyst reviews Form 4 for appropriate data elements and enters them on Form 6 as indicated. Comments on Form 6 cue the analyst to examine assumptions which need to be considered in interpreting the data.

The data elements on Form 6 are those needed to calculate the acquisition estimates and transfer of training formula (as shown in Chapter 3). In future developments this part of the model could be part of a computer subroutine. The analyst would enter the data elements and the computer will calculate all possible acquisition and transfer measures, note where there is insufficient data, and note when alternate sources yield similar or differing results (i.e., from a predecessor TD/S vs. a similar TD/S).

FORM 5: TASK ANALYSIS COMPARISON CHART

This form provides guidance for analyses to be used with new TD/S & Predecessor TD/S. One form is to completed for each comparison. Adapt the form to the particular categorization of tasks, subtasks, skills or excercises appropriate to both TD/S.

Name/Code of New TD/S: ____________________

Name/Code of Predecessor TD/S ____________________

Analyst's name ____________________ Date Completed __________

List task/subtasks/skills/exercises for New TD/S	List Tasks, Subtasks, Skills for Predecessor TD/S 1 1.1 1.2 1.3
1.	
1.1	
1.2	
1.3	
Subtotal Average	
2.	
2.1	
2.2	
2.3	
Subtotal Average	
3	
3.1	
3.2	
3.3	
Subtotal Average	
Total Average all predecessor tasks	

Code each task, subtask, skill as follows:

1. Identical-the task, subtask, skill or excercise is the same or almost the same as that in the comparison case.
2. Quite similar, but not identical; more than ½ similarity.
3. Similar - about ½ similarity.
4. Dissimilar - less than ½ similarity but not totally different.
5. Different - the tasks, subtask, skill or excercise is very dissimilar or entirely different from that in the comparison case.

FORM 6: SUMMARY OF BASELINE DATA AVAILABLE FOR ANALYSIS

Review the information sources on Form 4 and enter the data on this Form to determine the type and quality of data available. The baseline data should be helpful in guiding the design of the analytic methods. Complete one copy of this form for each WS and TP appropriate to the new TD/S and its courseware.

1. Analyst's name ______________________ Date completed ___________________

2. TD/S name & number ____

3. TP(s) name & number ____

4. WS(s) name & number ____

5. WS time in hours or trials to criterion allocated to training. (From Form 4, 4.6)

   ____

   ____

   ____

   ____

6. WS performance criterion measures. (Describe briefly from Form 4, 4.4)

   ____

   ____

   ____

   ____

7. TD/S time in hours or trials to criterion. (From Form 4, 5.9)

   ____

   ____

   ____

   ____

Comments

5&6 Reliability depends on whether this estimate is obtained from a training design or fielded training program, and the reliability of the criterion.

7&8 This information will evolve with the TD/S design, but should be specified early to aid in design iterations and forecasting.

8. TD/S performance criterion measures. (Describe briefly from Form 4, 5.5)

____

____

____

____

9. Predecessor TD/S

9.1 Time or trials to criterion in training (From Form 4, 6.2.4)____________________

9.2 Transfer of training data. Specify type of measure and result. (From Form 4, 6.2.6)

____

9.3 TD/S performance criterion measure and WS criterion measure. (Describe briefly from Form 4, 6.2.3, 4 and 6.)

____

____

____

____

9&10 Predecessor and similar TD/S should take account of similarities and dissimilarities to the new TD/S. The analyst may wish to adjust the data based on these judgments or to submit the data to a sample of SME's as part of the information to be used with an analytic method.

10. Similar TD/S. Data may be obtained from the data base or another source.

10.1 Time or trials to criterion in training (From Form 4, 7.5.4)

10.2 Transfer of training data. Specify type of measure and result. (From Form 4, 7.5.5)

10.3 TD/S performance criterion measure and WS criterion measure. (Describe briefly from Form 4, 7.5.3)

____

____

____

____

The cost analysis submodel may also be invoked at this point to examine cost implications of alternative designs.

From a TECIT research standpoint, the process is once again documented, leaving an audit trail of the sources of data employed, the input/output data of the baseline analyses, and the input provided to the TECIT analytic component.

DOCUMENTING THE CHARACTERISTICS AND EFFORT OF THE STUDY TEAM AND SUBJECT MATTER EXPERTS (SMEs) (Form 7)

Form 7 gives a method for documenting the characteristics, roles, responsibilities, background, experience and effort expended by the study team and the SMEs. The form is used to guide the analyst in selecting study team members and SMEs. Their selection will depend in part on the information gathered and the need for additional information as the design progresses.

Design and development of a TD/S calls for the assignment of a project manager and additional members of a project team who provide input and support to the TD/S design and development process. This team may be responsible for concept development, developing statements of work for contractors, and overseeing the TD/S through all phases of development, validation, production and deployment. The expertise and time to employ the transfer model may or may not be available among members of the TD/S team. Hence, it is useful to think of a separate, often overlapping, team specifically tasked to address the acquisition and transfer issues. As TD/S development team members are often too close to the problem, it is frequently advisable to obtain independent estimates of transfer and costs from other SMEs.

The forecasting study team is the team that designs the forecasting transfer project and assembles the information input required for the analysis. They may also choose to make their own forecasting estimates. However, in many cases, they will need assistance in identifying sources, study planning, making the forecasting estimates, analyzing the data and interpreting the results, tasking or contracting with additional SMEs to carry out these functions.

At present, this type of data is lacking in CTEA training development models (Goldberg and Khattri, 1986). As a result, there is little coherent knowledge about the types of people involved in TD/S design and forecasting and the effort expended in the analysis. If faithfully completed these data will provide better information by which to judge the cost and value of information.

From the point of view of research on forecasting there are also concerns about the reliability and validity with

FORM 7: DOCUMENTING THE CHARACTERISTICS OF THE STUDY TEAM AND THE SUBJECT MATTER EXPERTS (SME s)

Provide the data below to identify the roles, responsibilities background and experience of all members of the study team and SME s involved in the TECIT analysis. Make the entries as each individual is added to the project giving their name or ID number at the top of the form. For 1, 3, 4 check all that apply; for 2, give years of experience; 5 and 6 call for effort estimates in terms of man-hours expended or contractor costs. Complete one form for each course or WS. Additional forms should be used when there are more than 5 team members or SME s.

Analyst's name ______________________ Date completed ______________

TD/S name & number ____

WS name & number ____

Name, ID					
1. ROLE/RESPONSIBILITY (check)	1	2	3	4	5
1.1 Forecasting Transfer Team Leader					
1.2 Forecasting Transfer Team Member-Analyst					
1.3 Contractor					
1.4 Study Design and Analysis					
1.5 SME for Forecasting Estimates					
1.6 Other - specify					

FORM 7: (con't)

Name, ID	1	2	3	4	5
2. EXPERIENCE					
2.1 Total - Enter Years					
2.2 Experience - Transfer of Training-Enter Years 2.2.1 Transfer Research & Development 2.2.2 Practice in schools & job					
3. BACKGROUND - SPECIFIC TO SYSTEM - (Individual is knowledgable in: check)					
3.1 Weapon System 3.2 Training Related to WS 3.3 Predecessor TD/S or Training 3.4 Similar TD/S or Training					
4. BACKGROUND - EDUCATION AND EXPERIENCE (check)					
4.1 TD/S Development 4.2 Training Development 4.3 Education Technology 4.4 Military Instructor 4.5 Civilian Instructor 4.6 Psychologist-human factors 4.7 Psychologist-educational or cognitive					

FORM 7: (con't)

Name, ID	1	2	3	4	5
4.8 Engineer					
4.9 Operations Research					
4.10 Cost Analyst					
4.11 Economist					
4.12 Military					
4.13 Civilian					
4.14 Other (specify)					
5. MAN HOURS EXPENDED					
6. CONTRACTOR COSTS THIS TASK					

Comments. Note item number and individual name or ID.

which various SMEs make forecasting estimates. One part of a strategy for research on forecasting calls for maximizing SME variance along with other types of variance. Maximizing variance attributable to SMEs calls for adding independent judges so that the number is sufficiently large to be able to compare background and experience characteristics and to test for reliability. While additional SMEs add cost and effort to a forecasting study, a great deal may be learned that will in the future provide better guidance for their selection.

This form will also be included on the computer in future development of the model so that the analyst can be reminded to enter information at various iterations in the analysis and cumulative effort analyses can be made as team members and SMEs are added.

IS A TD/S NEEDED?

Form 8 gives a checklist for making a preliminary determination as to whether or not a TD/S is needed. An X in items 1-4 in the column shown indicates that a TD/S is needed. A 0 in items 1-4 in these columns indicates that a TD/S is not needed. Entries in the "not sure" column call for the development of one or more TD/S concepts for further analysis so that a definite yes or no can be given. The analysis should address the question of what tasks can be most cost effectively taught in conventional classroom instruction using training aids, the TD/S or the WS. If item 5 can be answered yes with assurance, it may serve as a "tie breaker" for a "not sure" in item 4. A yes in item 6 may break ties for a "not sure" in items 2 and 3.

SUMMARY

This chapter has given a detailed presentation of Component 1: Problem Definition of the Training Effectiveness Submodel of TECIT. The rationale, forms, applications, and research uses have been explained. By guiding the analyst through a set of cues and queries, the forms focus attention on information needed to:

(1) determine whether a TD/S is needed
(2) aid in designing an appropriate TD/S
(3) gather baseline data on acquisition and transfer of training
(4) provide an audit trail for applications and research
(5) show the context and purpose(s) for which analyses are made
(6) set the stage for designing analytic studies

System and non-system TD/S designs are considered. The uses and limitations of predecessor and similar TD/S are noted and incorporated in the analysis. Computerization of the model is discussed for future development. Documentation is provided of study team and SME characteristics and effort.

The next step is to design and execute analytic studies of acquisition, transfer of training, job readiness and safety. These methods are presented next in Chapter 3.

Form 8

IS A TD/S NEEDED?

YES	NOT SURE	NO	
X ---	---	O ---	1. Do safety and emergency procedures need to be practiced in a realistic setting before practicing on the WS or job itself or as refresher before resuming work on the job or WS?
X ---	---	O ---	2. Is practice required in integrating skills and knowledges in a realistic setting?
O ---	---	X ---	3. Can classroom instruction with conventional training aids provide realistic integration for all tasks?
X ---	---	O ---	4. Will a work sample of tasks and skills found on the job or in battle provide more realistic training and job (battle) readiness than can be provided during training by work on the WS or through conventional classroom instruction?
X ---	---	O ---	5. Are life cycle costs for a TD/S likely to be equal to or lower than training aids in classroom instruction?
X ---	---	O ---	6. Are life cycle costs for a TD/S likely to be lower than those on the WS?

Key:

X - TD/S needed for X in any one of 1-4. An X in 5 and 6 weighed in relation to 2-4.

O - TD/S not needed if 1 through 4 are all O.

Checks under "not sure" require development of TD/S concepts and further analyses to delineate benefits of tasks.

Chapter 3

TRAINING EFFECTIVENESS OF TECIT: ANALYTIC COMPONENT

INTRODUCTION

This chapter presents the TD/S function and its elements in further detail, showing how each element is obtained, weighted and used by the analyst. Analytic instruments for securing data are presented or referenced in each section. The discussion for each section considers analytic, empirical and research methods.

The chapter unfolds as follows:

1. The TD/S function is presented along with its elements and how they are obtained.

   1.1 Acquisition learning

   1.2 Safety and accident reduction

   1.3 In-course transfer of training - time to criterion measures and performance measures are discussed

   1.4 Job or battle readiness

   1.5 Utilization ratio and instructional management

   1.6 Weighting effectiveness elements

2. Diagnostic analyses are discussed. Two approaches are illustrated: task level diagnoses and diagnoses by estimating variance sources.

3. Cost effectiveness decision rules are discussed in brief.

4. Time, performance, safety, job readiness and cost trade-offs are discussed.

5. Multiple course uses of the model and exportability are discussed.

Chapter 2 dealt with problem definition and information gathering. Once the problem has been defined and background information obtained, the analyst is ready to proceed with the selection of the types of data appropriate to the TD/S. Next, the analyst identifies sources for additional information gathering by reviewing the types of information needed and the SMEs from which they may be obtained.

Finally, the analyst formulates interviews or questionnaires for use with SMEs to make the estimates.

The chapter demonstrates that the model is parsimonious. Only a limited number of transfer measures and data elements need to be considered for any given problem. If the number and types of transfer measures and data elements were very large, estimating them would be difficult.

Future computerization of the model would lead the analyst through a description of the formulae and queries regarding needed data elements. The analyst may then make preliminary estimates or begin the development of the questionnaires to obtain estimates from SMEs.

THE TD/S EFFECTIVENESS FUNCTION AND ITS ELEMENTS

As noted in Chapter 1, the TD/S effectiveness function is as follows:

$$\text{TD/S E (f)} = \left\{ \frac{\text{S, ToT, JR}}{\text{Acq}} \right\} \text{UR}$$

Where

TD/S E refers to the training effectiveness function.

Acq. is acquisition learning on the TD/S measured in terms of time to criterion on the TD/S.

S is a safety rating.

ToT is transfer of training from the TD/S to an exercise on the WS during training measured in various ways such as time savings or performance gains on the WS attributable to training on the TD/S.

JR is a rating of job readiness for a work sample TD/S, alternately defined as the transfer of training from the TD/S to the job, a battle exercise after training, or the skill maintenance retraining schedule required to maintain readiness.

UR is the utilization ratio of the TD/S defined as the hours used divided by the hours scheduled, times 100.

The analyst starts by selecting the appropriate elements and then turns to methods for estimating and weighting them. Acquisition and the utilization ratio are always included. Depending on the purposes and expectations of the TD/S, the

analyst selects one, two or all three of the safety, in-course transfer and job readiness elements.

ACQUISITION LEARNING ON THE TD/S

Acquisition on the TD/S is a necessary element in that judgments about safety, transfer of training and job readiness all impact time, performance and the criterion in TD/S acquisition. For example, if safety is a concern, there must be sufficient practice on the TD/S to assure that the trainee is ready to practice on the WS. Similarly, if a work sample TD/S is designed there must be sufficient practice on the TD/S to assure job or battle readiness. The same point applies to transfer of training to a WS exercise within the course.

Acquisition learning measures are also the first empirical data to be obtained when the TD/S is fielded.

The measures employed for acquisition learning include time, performance and a criterion of acceptable performance. However, these measures may be structured differently depending on how the TD/S is to be used in training. The three sets of measures are:

1. Variable time (trials, repetitions) - fixed criterion. The trainee takes as much time or as many trials or repetitions as needed on the TD/S to reach an established criterion. Averages of time, trials or repetitions are estimated. These types of measures have been used for flight training. They are appropriate when safety and emergency procedures are a concern and when it is important for the trainee to achieve the criterion on the TD/S before proceeding on to other training or to graduation from the course. Gunnery training is an example. To use these measures, the following conditions must apply:

    (a) A reliable performance criterion can be devised from task analyses or statements of objectives;

    (b) Variable time (trials, repetitions) requiring individual attention in the TD/S must be implementable in the training program. Of course, time is not infinitely variable, so a practical time limit may be imposed for the slowest trainees.

2. Variable performance, fixed time - Average performance is estimated. If a criterion is available, the percentage of the criterion may be

obtained or if the performance measure has a maximum score, the percentage of the maximum may be obtained. Fixed time sessions are often established when training on the TD/S requires substantial set up time, when teams rather than individuals are being trained, when training logistics tend to make it infeasible to train to criterion, or when criterion performance on the TD/S is not considered critical to safety or subsequent performance.

3. Variable time - variable performance. Used most often in empirical studies to find out how much time is needed to achieve various performance levels. Groups of trainees are given different time limits (or numbers of trials or repetitions) and average performance is estimated for each group. This approach is sometimes used to aid in establishing the performance criterion for the TD/S.

Only one set of measures should be used. The selection of the appropriate measures should correspond to those used in in-course transfer of training, when appropriate. Acquisition time and performance are also considered in relation to safety, job readiness and utilization. Task estimates or judgmental variance estimates may be made. Comparative analysis for acquisition may be made for two or more TD/S design alternatives by considering variations in time, performance levels or the criterion.

SAFETY AND ACCIDENT REDUCTION

Where safety is a primary concern, considerable time may be spent teaching emergency procedures on the TD/S prior to work on the WS because many tasks are too dangerous to do otherwise. Prime examples are in space flight and in the nuclear industry. Training is accomplished on the TD/S by simulation of all foreseeable contingencies before use of the actual equipment.

The sequence of instruction affects the transfer paradigm. For tasks that otherwise would be unsafe for trainees to perform, both the transfer and control group receive instruction first on the TD/S. The transfer group continues with instruction on safe tasks, followed by an exercise on the WS; the control group moves directly to the WS exercise. The sequence is summarized as follows:

TD/S Practice of Unsafe Tasks on WS	TD/S Practice Safe Tasks	WS Exercise
Control	--	Control
Transfer	Transfer	Transfer

The typical transfer experiment follows only the last two steps. In general, when practice on a TD/S is required before working on a WS because some tasks are unsafe, transfer estimates obtained underestimate true transfer values. The effect is quite direct on the Transfer Effectiveness Ratio (TER), increasing the magnitude of TD/S time to criterion. The effect is indirect on all transfer formulae as practice on unsafe tasks on the WS is likely to generalize to practice of safe tasks on the TD/S and the WS. These effects are confounded and there is no way that measurement methods can take them into account. The best solution is to separate the analysis into safe and unsafe tasks. This is only a partial solution as parts of some tasks fit in both categories, and generalization from unsafe to safe tasks is not taken into account. Therefore, more weight should be given to acquisition on the TD/S. This needs to be taken into account in the MAUM weighting given to safety.

Analytic judgment is required in designing the training for unsafe tasks and emergency procedures. As empirical accident data and experience accumulate, they are often incorporated into the TD/S courseware, however, accumulation of data and experience on specific WS requires many years of lead time. Some TD/S software is designed to be easily reprogrammed to take account of newly recognized hazards.

The analytic scale in Form 9 is presented for use with unsafe tasks and emergency procedures. The results of this analysis are entered on the summary profile form presented later in this chapter. All data is then reviewed and adjustments in time and criterion levels on the TD/S are made where appropriate. The training sequence is considered in terms of the amount of practice required on the TD/S prior to training on the WS. As experience accumulates the scale can be used in modifying the TD/S.

Reliability and validity of the estimates are very important. Accidents vary a great deal in terms of property damage, injury or death of personnel, costs, morale and public relations. Appropriate experts in TD/S designed for safety and accident reduction should be employed to make these judgments. Databases of accident reduction estimates are too broadly categorized to lend themselves to interpretation by persons less than expert in the safety field.

Form 9

Rating Scale for Safety and Emergency Procedures

To what extent is this TD/S (tasks, subtasks, exercise) expected to reduce the chances of an accident? In other words, to what extent are safety and accident reduction one of the purposes for which this TD/S (task, subtask, exercise) was (is being) designed? Rate the chances of reducing accidents as a result of training with this TD/S as follows:

0 - Not at all. Not a purpose of this TD/S or any of the tasks or exercises within it.

1 - Very low

2 or 3 - Low

4 or 5 - Average

6 or 7 - High

8 or 9 - Very high

If your rating was 1-9, rate the tasks, subtasks, or exercises using the scale of 1 to 9 above.

It should be noted that safety considerations may add to training time, performance criterion levels and costs. The empirical literature needs to document these relationships more fully.

As with the acquisition measures, the safety measures are weighted using MAUM methods (described later in this chapter) after all functional elements are considered. Comparative analyses of alternate concepts are conducted as before.

IN-COURSE TRANSFER OF TRAINING

General

After acquisition learning on the TD/S, in-course transfer of training is the next set of empirical data obtained after the TD/S is fielded. It is an appropriate measure when a relevant and reliable exercise on the WS is also included within the same course. In-course transfer is measured by comparing performance on the WS of a group that did (will) not receive instruction on the TD/S with one that did (will) receive instruction on the TD/S.

Transfer of training measures are classified in two ways:

1. Time to criterion measures of transfer of training.

2. Performance measures of transfer of training.

Time (Trials) to Criterion Measures of Transfer of Training

Time to criterion measures include the Transfer Effectiveness Ratio (TER) and the Percent Time Saved (PTS) on the WS. (PTS is sometimes called percent transfer or transfer ratio in the literature. We reserve the term percent transfer for performance transfer measures.) Both measures are "savings" measures of time on the WS. These measures were popularized by Povenmire and Roscoe (1971) and reviewed by Orlansky and String (1977, 1979, 1985) for applications to flight training. The Orlansky and String database on flight simulators is presented in summary form as part of Appendix A of this report and is a useful reference for comparison purposes.

These time- or trials-referenced measures are applicable to weapon systems other than aircraft when the appropriate assumptions can be met and when time or trials, given like

content on the WS and TD/S, are important variables in training and job measures. The key assumptions are the following:

1. Time or trials to criterion on the WS can be clearly specified and varied.

2. The criterion performance can be specified and is commonly agreed to.

3. The TD/S is not developed primarily for safety training or job readiness.

The time to criterion measures, in contrast to performance measures of transfer, have the advantage of using a common set of data elements - the common time metric. Performance measures are unique and specific to a particular area of application such as gunnery, maintenance, or tank commander training and sometimes to specific WS and levels in training. Performance measures are not as easily related to costs.

The time to criterion formulae, as with all transfer measures, are based on experimental paradigms that include experimental transfer groups, i.e., those using the TD/S, and control groups, those using only the WS. The experimental transfer paradigm is summarized as follows:

1. Control Group - WS time (trials) to criterion

2. Experimental Group - TD/S time (trials) to criterion; WS(TD/S) time (trials) to criterion

The formulae for these measures are as follows:

1. Transfer Effectiveness Ratio (TER) = $\dfrac{\text{WS} - \text{WS(TD/S)}}{\text{TD/S}}$

2. Percent Time Saved (PTS) on the WS =

$$\frac{\text{WS} - \text{WS(TD/S)}}{\text{WS}} \times 100$$

Common data elements on time to criterion measures are defined as follows:

WS = Time (trials) to criterion performance on a WS for a group that did not use the proposed TD/S. Represents the control group in an

experimental transfer design. Also represents the total training time for practice before a TD/S is introduced.

WS(TD/S) = Time (trials) to the same criterion on a WS for the group(s) using the TD/S. Represents one data item of the experimental group in an empirical transfer design. May be systematically varied or an average may be obtained in a given study.

TD/S = Time to criterion on the TD/S. Represents the "experimental treatment" in an empirical transfer study. The criterion on the TD/S may vary in its similarity to the criterion on the WS. In many cases it is quite similar and in other cases it is not. Time to criterion may be varied for a specific study to test the effects on WS(TD/S) time trade-offs. TD/S time to criterion may also be considered a measure of acquisition efficiency for the TD/S.

Time measures are expressed in terms of hours or fractions thereof. Trials or repetitions may be converted to average hours for costing purposes, but do not require conversion for empirical or analytic purposes.

As measures of transfer, they have the following characteristics in common:

1. No transfer occurs when WS = WS(TD/S). That is, no time savings have been achieved as a result of introducing the TD/S.

2. Transfer with negative effect occurs when WS is less than WS(TD/S). That is, it takes more time to reach criterion on the WS with the TD/S that without it. Presumably, learning the TD/S tasks interferes with learning the WS tasks. If found, the analyst should reexamine the TD/S design, the training curriculum, and the WS criterion.

Note that the numerators are identical in both formulae. The difference in the two formulae is in their denominators, with TER using TD/S time to criterion and PTS using WS time to criterion of the control group. By leaving TD/S time out of the formula, PTS fails to take account of acquisition learning on the TD/S. Rose and Wheaton's (1985) formulation of DEFT considers acquisition efficiency on the TD/S and also points out its importance in forecasting transfer and designing TD/S. This omission limits the usefulness of the PTS formula for designing TD/S, understanding the acquisition learning process, and relating an effectiveness measure to costs.

Differences in the results given by each of the two formulae are illustrated in Tables 13 and 14. Both tables show the common numerator WS - WS(TD/S) varying from 5 to 40 hours.

Only positive values are shown to indicate positive transfer. Table 13 shows TERs for various values of TD/S time to criterion. Note that TER values range a great deal depending on TD/S time, a measure of efficiency of the TD/S. Another way of expressing this result is as follows: for any given time savings on the WS, transfer as measured by the TER will vary with the time taken to acquire knowledges and skills on the TD/S.

Table 14 shows PTSs for various values of WS, the control group in the experiment. These values show that the PTS is relative to the amount of time originally required to reach criterion on the WS. Since as a practical matter it is important for soldiers to receive some amount of training on the WS, the job for which they are trained, PTSs that are too high, say 80% to 90%, may be substituting too much TD/S time for WS time. Thus, PTS as a measure has practical limits that can be best determined by a TD/S vs. WS tradeoff study.

Since the TER and PTS share the same numerator one might expect them to be highly correlated. However, Orlansky and String (1977, 1979) found that the two measures are correlated only $r = 0.49$ across a sample of 34 studies, accounting for only 24% of common variance. Hence, the different denominators in the TER and PTS contribute substantially to differences in results. TER takes account of learning time on the TD/S while PTS does not.

It should also be noted that negative values of TER will always yield negative values of PTS. That is, a TER cannot be negative while a PTS is positive. This is because the numerators in both formulae are identical, positive or negative, and the denominators in both formulae are always positive.

<u>The Truncated Transfer Effectiveness Ratio.</u> A truncated TER is one in which some students do not achieve criterion on the TD/S. The PTS has been used in empirical studies in some cases when there were training time constraints that prohibited all trainees from reaching criterion on the TD/S, reasoning that a TER would be misleading when the assumption for the TD/S group is not met. For analytic purposes, a truncated TER is recommended. It is not clear from the empirical literature whether a truncated or time limited estimate of TD/S time has been used in place of a TD/S time to criterion estimate in a TER formula. In empirical studies using the PTS, TD/S time has not usually been reported. This is a serious reporting deficiency in

Table 15

Transfer Effectiveness Ratio (TER) Function

Time to Criterion on the TD/S	Time Savings on the WS: WS - WS				
	5	10	20	30	40
100	.05	.10	.20	.30	.40
50	.10	.20	.40	.60	.80
40	.12	.25	.50	.75	1.00
20	.25	.50	1.00	1.50	2.00
10	.50	1.00	2.00	3.00	4.00
5	1.00	2.00	4.00	6.00	8.00

$$TER = \frac{WS - WS_{TD/S}}{TD/S}$$

Table 14

Percent Time Saved (PTS) Function

Time to Criterion for the Control Group: WS	Time savings on the WS: WS - WS				
	5	10	20	30	40
100	5	10	20	30	40
50	10	20	40	60	80
40	12.5	25	50	75	100
20	25	50	100	--	---
10	50	100	---	---	---
5	100	---	---	---	---

$$PTS = \frac{WS - WS_{TD/S}}{WS} \quad X\ 100$$

empirical studies. If PTS shows positive transfer, TER will also show positive transfer. Reporting the TD/S time, the time limit and the percent of students reaching criterion would enable readers to make their own estimates of TER, changes in training time resulting from adding the TD/S and cost-effectiveness analyses. For purposes of analytic forecasting the TECIT Model distinguishes between TD/S time to criterion and truncated or time-limited TD/S time. Both are useful in the TER formula but each calls for a different interpretation.

<u>Training ime Changes From Adding a TD/S to Training.</u> If there is positive transfer, introduction of a TD/S into the training environment will usually affect the total time required for training. The "new" time needed is TD/S + WS(TD/S), while the "old" time is WS.

A third measure is defined below to reflect the effect of adopting a TD/S on new total training time as follows:

Proportion Total Training Time Saved/Added (PTTS/A) =

$$1 + \frac{[(WS(TD/S) + TD/S)] - WS}{WS}$$

where all terms are as defined earlier.

Total training time is a matter of concern in planning and implementing training and needs to be considered along with transfer and costs. Even if the TD/S evidences transfer to the WS and costs less to operate than the WS, if total training time has to be increased substantially to implement the TD/S, there may be strong resistance to allocating additional training time. Perhaps this is the reason that some empirical transfer studies encounter the constraint of being unable to train to criterion on the TD/S and use the PTS formula instead of the TER formula. Use of truncated TD/S time will be useful in this case.

Fortunately, as shown in Table 15, there are parametric relationships among TER, PTS, and PTTS/A.

1. When TER = 1.0, total training time is unchanged; total training time with a TD/S is equal to total training time without a TD/S; or WS = WS(TD/S) + TD/S and PTTS/A = 1.0. PTS does not have any effect on PTTS/A.

2. When TER is less than 1.0 (0.2, 0.5, and 0.8 in Table 15), total training time increases. Reference to Table 15 shows that PTTS/A increases

Table 15

Proportion Total Training Time Saved/Added (PTTS/A) As A Function Of The Transfer Effectiveness Ratio (TER) And Percent Time Saved (PTS)

Percent Time Saved (PTS)	Transfer Effectiveness Ratio					
	0.2	0.5	0.8	1.0	2.0	3.0
10	1.4	1.1	1.02	1.0	0.95	0.93
20	1.8	1.2	1.05	1.0	0.90	0.87
40	2.6	1.4	1.10	1.0	0.80	0.73
50	3.0	1.5	1.12	1.0	0.75	0.67
60	3.4	1.6	1.15	1.0	0.70	0.53
90*	4.6	1.9	1.22	1.0	0.55	0.40
100*	5.0	2.0	1.25	1.0	0.50	0.33
	<-----Added **-->			—	<-Saved *->	

Formulae:

$$TER = \frac{WS - WS_{TD/S}}{TD/S}; \quad TD/S = \frac{WS - WS_{TD/S}}{TER}$$

$$PTS = \frac{WS - WS_{TD/S}}{WS} \times 100$$

$$PTTS/A = 1 + \frac{(WS_{TD/S} + TD/S) - WS}{WS}$$

* PTS's of 90% and 100% are shown to illustrate limiting values. They should be encountered infrequently as in most cases, the TD/S is not considered a complete replacement for WS practice.

** "New" time = "old" time on WS x PTTS/A

as PTS increases and TER decreases. For example, a TER of 0.2 and PTS of 60% yields a PTTS/A of 3.4, meaning that "new" training time is 3.4 times that required compared with "old" training time; however, a TER of 0.8 and PTS of 20% yields a PTTS/A of only 1.05, meaning that only 5% more "new" time is needed compared to "old" time.

3. When TER is greater than 1.0, there is a reduction in total training time depending on the magnitude of PTS. The larger the PTS, the more time is saved as measured by PTTS/A.

4. When TER is 0.5, the PTTS/A shows that total training time is added in the same proportion as it is saved on the WS as measured by PTS. Examination of the TER = 0.5 column in Table 15 shows that a PTS of 10% yields a PTTS/A of 1.1 or a 10% increase in total training time; a PTS of 50% yields a PTTS/A of 1.5 or a 50% increase in total training time.

The reader should bear in mind that TD/S life cycle costs relative to WS life cycle costs are important variables not yet considered. Time may be added to training by a TD/S to the extent that TD/S operating costs are less than WS operating costs. Cost effectiveness decision rules are discussed later in this Chapter and in Volume II.

Empirical and Parametric Time Measures Compared. It is instructive to compare Orlansky and String's (1977, 1979, 1985) empirical data on flight simulators in Appendix A with the parametric values in Table 13, 14, and 15. Orlansky and String's central tendency and variability statistics are summarized in Table 16. For flight simulators at the median (TER = 0.48, PTS = 41%), PTTS/A would be about 1.4 or 40% more time would be required. For flight simulators at the first quartile (TER = 0.20, PTS = 20%), PTTS/A would be about 1.8 or 80% more time required. Because of scatter in the TER/PTS relationship the third quartile and highest values cannot be directly interpreted in terms of PTTS/A. The lowest values in Orlansky and String's data were for the same case and showed negative transfer. Suffice it to say that in about one fourth of the sample, only a relatively small amount of training time was added (perhaps 5% to 20%), there was no change, or there was a reduction in total training time. For about three-fourths of the sample, more than 20% training time appears to have been added.

Thus, the cost-effectiveness of the majority of cases in Orlansky and String's database depends on a favorable ratio of TD/S hourly costs to WS hourly costs. This ratio averaged about .08, more than compensating in most cases for the increase in total training time. (It is not known

Table 16

Summary Of Orlansky And String's Transfer Of

Training Data On Flight Simulators: Central Tendency And Variability

Statistic	Transfer Effectiveness Ratio	Percent Time Saved
Median	0.48	41
Q_1 - First Quartile	0.20	20
Q_3 - Third Quartile	0.75	48
Highest Value	1.90	90
Lowest Value	-0.49	-10

Source: Orlansky and String (1977, 1979, 1985)

whether a truncated TD/S time was used in the studies reviewed. Individual study results may be reviewed by referring to Appendix A, Figure 5.)

Relative Downtime. The relative availability and convenience of use of a TD/S vs. a WS also need to be considered. It is useful to compare the availability of the TD/S and WS by an estimate of "downtime" derived from reliability/maintainability data. Although the relative reliability/ maintainability costs may be reflected in the cost analyses, we are concerned here with training availability, scheduling and reducing disruptions. The TD/S may be viewed as more useful for training to the extent that the downtime rate is more favorable than the WS. For example, assume the analysis suggests that 50 hours of training is needed on the TD/S and 60 hours of training is needed on the WS. Assume a downtime rate (hours inoperable/total hours) of 0.1 for the TD/S and 0.2 for the WS.

50 hrs. x 0.1 = 5 down-time hours lost on TD/S
60 hrs. x 0.2 = 12 down-time hours lost on WS

Given the number of students to be trained, the scheduling of training may be taken into account by adjusting each term in the TER, PTS, and PTTS/A formula, and corrected transfer and training time estimates obtained. In addition, consideration may be given to the number of spares and the spare parts requirements needed to maximize time in operation.

In the design phase, alternative TD/S concepts may be evaluated in terms of assumptions about reliability, maintainability and scheduling. In the fielding phase of the TD/S, reliability and maintainability data may be obtained and factored into the time and cost measures. Downtime is also considered as a part of the utilization ratio later in this chapter.

Discussion of Time to Criterion and Related Measures. The TER is recommended over the PTS formula for both empirical and analytic purposes, with distinctions made in terms of TD/S time to criterion vs. truncated or time-limited TD/S time. In the analytic mode, the untruncated TER should be obtained first and analyzed in relation to PTTS/A for total training time implications. If total training time is expected to be too large, truncated time may be analyzed in terms of its learning implications for trainees who do not reach criterion and for the cost-effectiveness of the truncated and untruncated TER.

Empirical studies of the truncated and untruncated TER in relation to limits of the utilization of training time, transfer for trainees who do not reach criterion on the

TD/S, and cost-effectiveness would be useful but have not appeared in the literature. Analytic studies could yield insights into how much difference a truncated TER would be likely to make. The practice of using PTS when TD/S time is truncated appears to have been justified on the basis of practical training time limits imposed by school personnel. However, if truncating the TD/S time still yields positive transfer it is easy to see that it reduces total training time and costs. Downtime estimates should be taken into account by estimating relative reliability and maintainability.

Performance Measurement and the Criterion

Transfer of training has often been measured in terms of performance alone. The performance measures and criterion should be defined early in the design of a TD/S. The criterion is the point on a performance measure(s) at which a trainee is classified as a "GO" or "NO-GO." The criterion is established in relation to analytically-derived training objectives (for the whole course or individual tasks) and measures of performance and time devised to measure these objectives. The training objectives themselves are derived from analysis of the threat scenario, performance requirements on the WS, and that part of the job for which training is being devised. Thus, in the TD/S design phase, analytic as opposed to empirical methods are used to define both the performance measures and the criterion of minimum acceptable performance. Empirical analysis is possible only when the measures have been operationalized for a predecessor or similar TD/S or the TD/S has been fielded.

Generally, there is greater complexity in performance measures as opposed to time to criterion measures. The type and number of performance measures may vary. There is sometimes one measure of performance, several measures treated separately, several measures combined, or measures appropriate to some tasks or skills but not to others.

The types of performance measures have been classified in two ways: knowledge vs. skill performance measures. Examples of skill performance measures include hits on target in gunnery practice or navigating to a correct location. The quality, amount and a time limit may be included as part of the performance criterion. For example: the trainee will make a minimum of 20 hits within a three-minute period within a range of two feet of the target center.

Objective measures are usually gathered through recording devices of some sort, but may require post-record keeping analysis to obtain final measurements. Subjective measures generally employ observation checklists and rating scales. Examples include such things as checklists for use of correct procedures in repairing a motor or making correct

maneuvers. As with objective measures, subjective measures may also include observations of quality, amount and a time limit.

The scales of measurement for objective measures are usually interval or ratio scales while those for subjective measures are usually ordinal scales. In contrast, time to criterion has the advantage of being a ratio scale.

Reliability of the measures is critical as without adequate reliability, variance on the WS measure may be too large to be able to detect true differences that may be attributable to the TD/S. Reliability is often measured by correlation coefficients, a procedure limited in its interpretation. More important for training purposes is accuracy at the criterion point of the performance measure. A discrepancy measure from the criterion is preferred in establishing the "spread" in relation to criterion performance. The standard error of the mean or median may be used and a percent discrepancy from criterion may also be employed.

The reliability of both the TD/S and the criterion on the WS are of equal concern. Many TD/S designs include in them automated or improved methods of scoring from which reliability estimates may be obtained. If the TD/S may also be considered a work sample for job readiness, then reliability on the TD/S may be used as a proxy for the reliability of the criterion. Otherwise, criterion reliability on both the TD/S and WS are needed. In many cases, reliability on a WS is enhanced only for research purposes as when photographic methods and additional observers are used in tank gunnery field exercises.

Performance Transfer of Training Formulae

There are many formulae for performance transfer of training. Three formulae are presented here for selection by the analyst.

The first formula is offered to take account of the criterion:

Percent Transfer to Criterion (PTC) =

$$\frac{T}{\text{Crit.}} \times 100 - \frac{C}{\text{Crit.}} \times 100 = \frac{T - C}{\text{Crit.}} \times 100 \qquad (1)$$

Where T and C scores reflect the average performance of a transfer and control group and Crit. is the designated performance criterion value for the measure in question. In all performance transfer formulae, T-C is the numerator when

a higher score indicates better performance than does a lower score. T and C are reversed when a lower score indicates better performance, such as in error measurements. In that case:

$$PTC = \frac{C}{\text{Crit.}} \times 100 - \frac{T}{\text{Crit.}} \times 100 = \frac{C-T}{\text{Crit.}} \times 100 \qquad (1a)$$

This formula was devised by the author to overcome shortcomings in the performance transfer measures currently found in the literature. When a high score means better performance, it has the following characteristics:

1. The components $\frac{T}{\text{Crit.}} \times 100$ and $\frac{C}{\text{Crit.}} \times 100$ will be equal to or exceed 100% when each group reaches or exceeds criterion performance. Otherwise, both components will be less than 100%.

2. The difference $\frac{T-C}{\text{Crit.}} \times 100$ gives a measure of transfer that is a constant for any given T - C difference relative to the criterion scale of measurement. For example, if T = 8, C = 6, Crit.= 10, then PTC = 20%. Similarly, if T = 9, C = 7 and Crit. = 10, the result is the same, PTC = 20%.

3. As it is the only performance transfer measure that incorporates the criterion level within it, it may be used in conjunction with time to criterion measures in empirical studies when both time and performance are jointly varied; in studies where performance to criterion is supposed to be constant, but still may differ between the T and C group; and in analytic studies where the design problem conceptually must consider the efficiency of learning by alternate designs, i.e., time to criterion and performance relative to criterion. For example, if two TD/S designs were both expected to yield a TER of 0.8, but the comparison expected to yield a PTC of 80% for simulator 1 and 100% for simulator 2, the second would be preferred.

4. When used in conjunction with training time and costs, the basis is established for explicit tradeoffs. For example, the preferred TD/S design is that which optimizes

the following: a high PTC, $\frac{T}{\text{Crit.}} \times 100$ of 100% or more, leaves training time unchanged or does not add a significant amount, and has costs favorable to the TD/S compared with the WS.

This formula is less easily interpreted when a lower score means better performance, as in error measurement or a time limit. In that case, transfer values over 100% indicate lower transfer or less than criterion performance, 100% indicates performance at criterion and values below 100% indicate higher transfer. The formulae is also subject to restriction of range as criterion errors or time approach zero. Two examples illustrate these points:

Example 1: Criterion = no more than 10 errors. Assume C = 15 errors, T = 5 errors. Then PTC = (15/10)(100) - (5/10) (100)=150% - 50% = 100%.

Example 2: Criterion = no more than 50 errors. Assume C = 40 errors, T = 20 errors. Then PTC =(40/50)(100) - (20/50) (100) = 80% - 40% = 40%.

If possible, scoring should be set up so that higher scores mean positive transfer to avoid this problem.

The second formula:

$$\text{Transfer Ratio (TR)} = \frac{T-C}{T+C} \times 100 \qquad (2)$$

has the advantage of limiting the range from -100% for negative transfer to +100% for positive transfer with zero equalling no transfer. However, it has the undesirable bias of yielding a higher PT when both T and C groups score low and a lower PT when both groups score high, presumably closer to the criterion. For example, when T = 8 and C = 6, PT = 14%; but when T = 15 and C = 13, PT = 7%. A scale with opposite characteristics would be preferred -- low transfer when both groups score low and high transfer when both groups score high.

It is a useful formula when low scores mean better performance i.e., errors or a time limit. In that case the formula becomes

$$TR = \frac{C - T}{C + T} \times 100 \qquad (2a)$$

As stated earlier the scale yields a higher TR when both groups score low, presumably closer to the criterion.

The third formula:

$$\text{Percent Transfer Max. (PTM)} = \frac{T-C}{Max} \times 100 \qquad (3)$$

where T and C are as before and Max. is the maximum possible score, assumes that there is a maximum for the measure in question, a condition that sometimes does not apply. However, if a maximum score can be designated, this formula has the advantages of ranging from -100% to +100%, and giving equal weight to equal T-C differences at all points on the scale. It has the disadvantage of not being criterion referenced. It is not useful for error or time measures where the minimum is zero.

A number of other performance transfer measures commonly found in the empirical transfer literature are:

$$\text{Percent Transfer} = \frac{T-C}{C} \times 100 \qquad (4)$$

$$\text{Percent Transfer} = \frac{T-C}{Max-C} \times 100 \qquad (5)$$

where T and C are as defined earlier and Max-C is the maximum score found in the control group sample.

The problems with these formulae are as follows:

1. Formula 4 has no definable bounds. It can range from ± infinity, making it specific to the particular sample in question and difficult to interpret. For example, if the T group averaged 50 hits on target in a gunnery exercise and the C group averaged 30 hits, then PT = ((50-30)/30)) x 100 = 67%. However if the C group averaged only 10 hits, PT = ((50-10)/10)) x 100 = 400%. Data cannot be compared from one study to another and the scale does not make any pretext of having interpretable intervals. It is particularly susceptible to error variance in the control

group.

2. Formula 5 is also susceptible to error variance in the control group, the particular level achieved by the control group and, in addition, the unreliability of the maximum score achieved in the control group sample. It too does not yield a scale with equal appearing intervals of measurement.

3. Neither formula takes into account the criterion performance level.

Formulae 4 and 5 are not recommended.

Discussion of Performance Transfer Formulae. The PTC formula (formula 1, including its components) is preferred and should be used whenever the criterion performance level can be designated and high scores mean better performance. If the criterion level itself is the subject of exploratory research or cannot be designated for some reason, then PTM (formula 3) may be used if the measure has a maximum score and high scores mean better performance. The TR = (T-C)/(T+C) x 100 (formula 2) may be used to limit values to a scale of $\pm$ 100% and is particularly useful when low scores mean better performance.

If multiple performance measures are used, they should be considered individually or weighted analytically in accordance with their worth in preparing trainees for the job.

Evaluating performance transfer requires intelligent examination of individual data items bearing in mind that the PT are summary measures. The following data items should be examined in udging the value of transfer:

1. Performance measure characteristics, range of values, reliability of TD/S and WS measures

2. The criterion value

3. Transfer group average estimate

4. Control group average estimate

5. Percent of criterion or maximum for the transfer group

6. Percent of criterion or maximum for the control group

7. Percent transfer to criterion, maximum or other formula

Caution should be used in interpreting performance transfer as a limited scoring range, 10 or 20 for example, may yield wide swings in the data.

Performance formulae for transfer of training do not take into account acquisition or acquisition efficiency on the TD/S. This is an important matter in the design of TD/S but has been virtually ignored in the empirical literature. It cannot, however, be ignored in TD/S design as acquisition learning is an important initial criterion in the efficacy of a TD/S. Rose and Wheaton (1985) in their formulation of the Device Effectiveness Forecasting Technique (DEFT) give major attention to the TD/S acquisition process as well as the transfer process. For an overview of this issue review the TECIT and DEFT conceptual framework and measures in Chapter 1 and the questionnaires in Appendix B.

It is unfortunate that many empirical studies have not used the transfer measures recommended here. To be most useful a review of empirical performance transfer measures would have to include data on the criterion and maximum score in order to recalculate the data to common transfer metrics. A separate study would be needed to find out if these types of data are available. It would be informative to analyze and compare the empirical distributions of the various PT formula for a variety of studies and compare their output and interpretability in conjunction with the PTC, training time measures and life cycle costs.

The reader is reminded that acquisition, training time and costs need to be taken into account for a full assessment of a TD/S.

Training Time Changes and Performance Transfer. The training time impact of introducing a TD/S also needs to be considered when performance measures of transfer of training are employed. The PTTS/A formula may be used to measure the restructuring of training time when performance measures of transfer are of major interest, however, the terms WS, WS(TD/S), and TD/S are redefined as fixed time or time limits rather than time to criterion. If there is no intent to save time on the WS, then WS = WS(TD/S) and total training time is increased by the time needed for the TD/S. However, time restructuring may also lead to a reduction of WS time, even if WS time savings is not considered a primary measure of transfer of training in the specific instance.

In contrast to the time-to-criterion transfer measures, there is no parametric relationship of PTTS/A and performance measures of transfer of training. Thus both measures have to be estimated independently. The relationship of practice time and performance on the TD/S

and the WS is an issue that has to be assessed analytically and empirically for the individual TD/S application.

Limitations of the Transfer of Training Paradigm

There is litle question that transfer of training is appealing in concept and useful in practice. However, it has a number of limitations which do not make it a completely satisfactory measure of the worth of a TD/S.

1. Transfer of training formulae are summary measures of data and as noted throughout this discussion may be unreliable or misleading. They should be considered in relation to the data elements from which they are derived. Also, only the Transfer Effectiveness Ratio includes acquisition learning on the TD/S as part of its formula.

2. Consideration of safety and hazardous conditions precludes the use of the traditional empirical transfer experiment. To avoid accidents, training on the TD/S is needed before training on the WS. Transfer as an analytic concept is still valid, but is not measured by the empirical transfer experiment. An analytic rating scale is presented earlier in this chapter for use in conjunction with analytic and empirical acquisition and transfer data.

3. TD/S are often developed to reconfigure training. (See definitions in Chapter 1.) That is, the training program and exercises on the WS may be modified at the same time as a TD/S is being developed to improve the instructional sequence, integrate knowledges and skills, provide realistic practice, improve instructional management, and improve the reliability and validity of measurement. They may also be designed to provide a better work sample of the knowledges and skills used on the job. The exercise on the WS may not be as good a work sample measure as the TD/S. Once again, the empirical transfer experiment may be limited in its application, particularly when a WS exercise in the training program does not provide a reliable or valid criterion against which to measure the effectiveness of the TD/S. Improvements in the instructional sequence and the intergration of knowledge and skills through realistic practice can be assessed by comparative analysis of acquisition learning (analytic or empirical), improved reliability and validity of measurement by commonly available measurement methods, and instructional management.

4. Job readiness is often used interchangeably in concept with transfer of training, but rarely if ever in empirical measurement. Job readiness refers to how far a course of instruction carries trainees toward being able to do the job proficiently. Given alternative TD/S and WS exercises, the better ones are those that carry the trainee closer to job proficiency. In a career sequence, degrees of job readiness are implied by training and work sequences such as basic and advanced training. Greater proficiency is expected at each level. For WS operator personnel such as tank commanders and gunners, it is virtually synonymous with battle readiness, while for maintenance and supply personnel both peace-time and mobilization readiness are considered. Work sample or criterion-referenced TD/S are designed specifically to include a wide array of important job and battle conditions and should demonstrate greater job readiness than other types of TD/S. Conceptually, transfer is implied. However, empirical transfer or follow-up studies from the end of a course to the job, or between courses in a career sequence, are difficult and costly to conduct and are rare in the literature.

JOB OR BATTLE READINESS AND WORK SAMPLE TD/S

Work sample or criterion-referenced TD/S are those devised to sample and replicate tasks or skills that are important, life threatening, may be infrequently encountered on the job or in battle, and are often not feasible to present on a WS exercise during training. Like a WS exercise, they provide practice in integrating skills and knowledges. For example, maintenance trainers are often designed to simulate faults infrequently encountered on the job; tactical simulators for operator personnel such as tank commanders may include battle conditions infrequently encountered even in the most realistic post-training field exercise. Examples of work sample TD/S in Armor School training include the Unit Conduct of Fire Trainer (UCOFT) and Simulated Combined Army Training (SIMCAT).

The TD/S and the WS exercise used in the training program are both work samples which are expected to contribute to job or battle readiness. Transfer of training from the TD/S to the WS during training can only reveal their common variance. The remaining unique variance of each one is that which each contributes to job or battle readiness.

An approach is needed that goes beyond that which can be empirically measured in the operation of the training program itself. The need for measures of job readiness is fairly straightforward. First, given two alternative work sample TD/S concepts, the preferred concept, other factors equal (i.e., career sequence, costs), is the one that carries trainees the furthest toward job or battle readiness. Second, given a work sample TD/S ready to be fielded, a judgmental estimate of improvement of job or battle readiness attributable to the TD/S: (a) can serve as a useful analytic tool, along with in-training empirical data to convey the worth of the TD/S; and (b) can be used to aid in determining the need for skill maintenance (refresher) training using the TD/S.

A job (battle) readiness judgmental scale may serve as an interim measure until empirical data may be obtained relating performance on the work sample to job performance or as an independent criterion of worth of the TD/S. Although transfer or job follow-up studies are possible, they have rarely been conducted and data from them would not mature for some time after the TD/S has been fielded. By way of example, examine Orlansky and String's (1977, 1979, 1985) data on maintenance and flight simulators shown in Appendix A. They summarize acquisition learning comparisons (time and performance) for maintenance simulators and point out the need for follow-up data. The flight simulator transfer studies reviewed use an "in-training" measure of performance on the WS, i.e., the amount of time saved on the WS in the course. Follow-up transfer measures were not used.

Form 10 shows a questionnaire for structuring the analysis. The form is used at the level of the TD/S as a whole or for tasks or task clusters. Scoring is shown at the end of the form.

Question 3.1 addresses the issue of skill decay and retraining, one of the conditions necessary for maintenance of battle readiness. The curve provided by these estimates may be used along with other data to plan skill maintenance training and to plan follow-up studies of field experiments of the TD/S and battle exercises. The form was also devised to obtain rater reliability and variance estimates in a manner similar to that used by Pfeiffer and his associates (1985) in FORTE.

UTILIZATION RATIO AND INSTRUCTIONAL MANAGEMENT

As noted in Chapter 1 and in Goldberg and Khattri (1986, Chapter 8) instructional management variances are expected to be related to the acceptability of a design and to utilization of the TD/S. Although no firm evidence is available to relate the two, Orlansky and String (1977,

Form 10

Job and Battle Readiness Questionnaire for Training Devices and Simulators

This questionnaire is for use by TD/S analysts, experts and contractors. Other personnel may be consulted as needed.

Work sample or criterion-referenced TD/S are those devised to sample tasks and skills found in the job or in battle that are important, infrequently encountered and may not be feasible to present fully on a WS exercise during training. Like a WS exercise they provide practice in applications and integrating skills and knowledges. Answer the following questions related to job (battle readiness) and use of the TD/S in training.

1. Estimate the percentage weight to be given to performance on the TD/S in determining whether an average student receives a GO or NO-GO for the course. Consult with senior schooh staff.

   0% - No weight. Not considered

   100% - All. The only consideration

   ____%

   1.1 Estimate 1 by trainee ability

   ____% 1.1.1 More able trainee

   ____% 1.1.2 Slower trainee

2. Consider how far the trainee has to go to be job ready once he/she completes training on the TD/S and WS. Use the following scale:

   0% - Completely job ready

   100% - Not ready. Further experience required

   ____% 2.1 Estimate as a percentage how far the average trainee has to go after completing the training program including all WS exercises.

   2.2 Estimate 2.1 by trainee ability

   ____% 2.2.1 More able trainee

   ____% 2.2.2 Slower trainee

   ____% 2.3 Estimate as a percentage how far the average

trainee has to go after completing the TD/S, but not the WS exercise.

2.4 Estimate 2.3 by trainee ability

____% 2.4.1 More able trainee

____% 2.4.2 Slower trainee

3. Consider a battle or mobilization exercise after training as a method of measuring job readiness. For each of the following, estimate the percentage contribution of the TD/S to battle or mobilization readiness for an average trainee. In other words, how much difference would training on the TD/S make between a group trained with the TD/S and a group not trained with the TD/S? (Assume that the battle mobilization exercise can be scored reliably to detect differences and that it is relevant to the TD/S.) Estimate percentages as follows:

0% - No contribution

100% - All. Complete contribution to readiness

3.1 If the battle or mobilization exercise was held within the following time periods, considering skill decay

____% 3.1.1 within 3 months of the completion of training

____% 3.1.2 within 6 months

____% 3.1.3 within 9 months

____% 3.1.4 within 12 months

____% 3.1.5 within 15 months

____% 3.1.6 15 months or more after training

3.2 Estimate the contribution of the TD/S to battle or mobilization readiness for trainees of differing abilities.

3.2.1 within 3 months

____% 3.2.1.1 More able trainee

____% 3.2.1.2 Slower trainee

3.2.2 within 6 months

____% 3.2.2.1 More able trainee

____% 3.2.2.2 Slower trainee

3.2.3 Within 9 months

____% 3.2.3.1 More able trainee

____% 3.2.3.2 Slower trainee

3.2.4 Within 12 months

____% 3.2.4.1 More able trainee

____%3.2.4.2 Slower trainee

Scoring:

Item 1: Weight in course.

Average % Rating

Item 2: How far the trainee has to go after training to be job ready. Low score indicates higher job readiness.

100 - Average % Rating

Item 3: Contribution of TD/S to readiness.

Average % Rating of 3.1.1 to 3.1.4

Overall Job Battle Readiness

Average of scores from items 1 - 3.

1979), Blaiwes and Regan (1986) and many others have commented on the problem of under-utilized TD/S. The implications of underutilization should be clear. Even if empirical studies demonstrate unequivocally that a TD/S contributes to transfer, safety, job readiness and cost savings in the experimental environment, these benefits will not be realized if the TD/S is not used, i.e., all values drop to zero.

The scale in Form 11 was devised to address the problem. It can be used at any phase of development of the TD/S, but should be addressed as early as possible. Items 1-8 provide a useful checklist for comparing alternate designs and contractor proposals, monitoring contractor development, and planning the implementation of the TD/S. The ratings in item 9 focus the analyst's attention on other variances that may affect utilization rates. It may not be possible, nor is it necessary, to give firm answers to each item at one time. The intent of the scale is to focus TD/S project managers' and contractors' attention on the problems that need to be addressed.

The scale is used at the level of the entire TD/S rather than at a task level. When alternative TD/S concepts are being considered, it could be used as a final review of the alternatives for final decision making. If no estimate can be obtained, a dummy variable of 100 should be used temporarily, assuming the TD/S will be used all of the time scheduled.

Research would be useful to establish the validity of the scale. A comparison of highly utilized and underutilized TD/S would be informative. Integrity of utilization is not considered here, but should be after the TD/S is fielded. This concept is also related to the general concept of technology transfer and in further developments might be integrated within that framework.

WEIGHTING EFFECTIVENESS ELEMENTS

Although each effectiveness element (i.e., acquisition, safety, in-course transfer, job readiness) may be evaluated separately, it may be useful to obtain a weighted combination of effectiveness elements to compare alternative designs or to obtain a summary measure of effectiveness when more than one element is applicable to a particular TD/S. The weighting method then yields a measure of the overall perceived value of the effectiveness of the TD/S. The weights are needed because the metrics for each element are not expressed in the same terms.

The weighting method uses a Multi-Attribute Utility Assessment Method (MAUM) similar to that used by Dawdy and Hawley (1982). The analyst examines the estimates for each

Form 11

Utilization Ratio - Instructional Management Scale

This form is for use by TD/S developers.

1-8. Have each of the following been adequately considered in regard to the concept and design of the TD/S for the course in question?

Yes	No	
---	---	1. TD/S (or alternatives) practice time, WS practice time, sequencing and scheduling.
---	---	2. Instructor/trainee ratio for the TD/S.
---	---	3. Instructor/trainee ratio for the WS.
---	---	4. Downtime for the TD/S (or alternatives) based on estimated reliability and maintainability.
---	---	5. Downtime for the WS based on estimated reliability and maintainability and ceremonial or other non-training uses.
---	---	6. Design of the instructor station for ease of use and operation, including such matters as selection of tasks, providing cues and feedback to the trainee, and scoring performance.
---	---	7. Instructor training for utilization of the TD/S.
---	---	8. Expert instructor input for items above.

For each item above answered "NO" further analysis should be considered.

9. Estimate the utilization rate (time used/time scheduled) x 100. Use the following scale:

0 - No use at all

100 - used all of the time scheduled

Rate probable utilization under each of the following conditions for the environment in which it is to be used. Enter scale values in spaces to the left.

9.1 School staff and instructor acceptance based on items 1-8 above.

____ 9.1.1 High

____ 9.1.2 Average

____ 9.1.3 Low

9.2 Command emphasis:

____ 9.2.1 Required

____ 9.2.2 Supportive, but not required

____ 9.2.3 Neutral

____ 9.2.4 Not supportive

9.3 School staff and instructor acceptance based on perceived face validity and data for acquisition, transfer, accident reduction, and job readiness.

____ 9.3.1 High

____ 9.3.2 Average

____ 9.3.3 Low

9.4 Considering your responses in 9.1 through 9.3, rate the probable utilization rate that is:

____ 9.4.1 Average or most likely

____ 9.4.2 Highest expected

____ 9.4.3 Lowest expected

Identify potential problem areas and address them.

data element and rates them with regard to importance and criticality for training and for job performance. The estimates of data elements in the design phase are analytic estimates. Empirical data should be substituted as it matures in the fielding phase, particularly for acquisition learning and in-course transfer.

Task level estimates (grouped or sampled) should be used whenever possible, particularly in the design phase, and summed over tasks. The MAUM effectiveness value of each task may then be examined and considered for inclusion or exclusion in the TD/S.

When more than one performance measure of transfer is used, the performance measures should also be weighted using similar procedures to obtain a single measure. The separate measures may also be retained for analysis.

Although the TD/S analyst and design team may make their own estimates, officers and expert instructor SMEs should also be employed to obtain a user perspective.

Methods and formula for the MAUM technique may be adapted from Dawdy and Hawley (1982).

SUMMARY PROFILE AND DIAGNOSTIC ANALYSIS

When the analyst has selected the primary measure(s) of transfer, and safety and job readiness have been considered, a detailed diagnostic analysis is in order. The set of data items and formulae are listed in detail on Form 12 for time to criterion as a primary measure and Form 13 for performance as a primary measure. Both forms list acquisition, transfer, the safety rating, the job readiness rating and life cycle costs. While a single overall analysis might be made, a diagnostic analysis would be more helpful. The diagnostic analysis may be made at the task or subtask level when such information is available or by using the sources of variance concept explained in Chapter 1 and illustrated throughout this chapter and in Appendix B.

Diagnostic task level analysis uses transfer, acquisition, safety and job readiness data (analytic or empirical) subdivided as far as practical into tasks, subtasks or skill elements. If the number of task elements is too large, tasks they may be grouped or sampled. This may or may not be possible for some simulators, but should be done whenever the WS tasks and TD/S tasks can be delineated. This disaggregation of the task elements yields a profile with all possible acquisition, transfer, safety and job readiness data. An empirical illustration is shown in Holman's profile of TERs in Appendix A. It should be noted that it is often not possible to obtain the same level of detail with empirical data as with analytic data.

Form 12

Illustration of Course Analysis Summary Diagnostic Profile When Time or Trials to Criterion are the Primary Measures of Transfer

Data Element and Formulae	Overall Course	Task Analysis, Variance Sources or Comparison of Alternative Concepts
1. Safety - Accident Reduction Rating		
2. WS-Control group time to criterion		
3. WS(TD/S)-Transfer group time to criterion		
4. TD/S-time to criterion		
5. TER (or truncated)-Transfer Effectiveness Ratio		
6. PTS-Percent Time Saved		
7. PTTS/A-Prop. Total Training Time Saved/Added		
8. Job Readiness Ratings		
9. Utilization Ratio*		
10. Operating Cost Ratio		

*For course as a whole only

Form 13

Illustration of Course Analysis Summary Diagnostic Profile When Performance Measures are the Primary Measures of Transfer

Data Element and Formulae	Overall Course	Task Analysis, Variance Sources or Comparison of Alternative Concept
1. Safety - Accident Reduction Rating		
2. T - Transfer group average on WS		
3. C-Control group average on WS		
4. Scale direction indicator- H or L. High score means better performance or Low score means better performance		
5. Crit-Criterion value on WS*		
6. Max-Maximum Score Value on WS*		
7. PTC-Percent Transfer to Crit.**		
8. PTM-Percent Transfer Max.**		
9. PT-Percent Transfer** = (T-C)/(T+C) (100)		
10. Time on WS - T group		
11. Time on WS - C group		
12. T group time on the TD/S		
13. PTTS/A: Porportion Total Training Time Saved/Added		
14. Job Readiness Ratings		
15. Utilization Ratio***		
16. Operating Cost Ratio		

*Depending on availability. Max. used only when high score means better performance.

**Selected according to availability of 5 or 6.

***For course as a whole only

Judgmental variance sources may be more useful, particularly if a detailed task list is not available and other variances need to be considered. The analyst may develop estimates of task complexity, task difficulty, criterion reliability (e.g., instructor leniency), student ability, physical fidelity and functional fidelity.

The questions that can be addressed to this diagnostic profile illustrate the value of the disaggregation for diagnostic purposes particularly in the TD/S design (or redesign) phase.

1. Based upon an examination of the training program design and TD/S design, are there any tasks, sub-tasks or skills that can be taught by some other training method or medium that would be likely to be more effective or equally effective and less costly?

2. Based upon an examination of the task profile of the TD/S for acquisition and transfer, particularly those tasks with low transfer results:

   2.1 are there ways to improve time to criterion and/or performance on the TD/S? If PTTS/A is too large, would a time-limited TD/S exercise be as likely to be as effective as one that allows as much time (trials) as the trainee needs?

   2.2 are there ways to improve transfer in terms of time reduction or performance improvement on the weapon system or the job?

   2.3 what are the cost/effectiveness implications of the alternatives considered under 2.1 and 2.2?

   2.4 are there ways to redesign the TD/S to reduce costs while maintaining an acceptable level of acquisition efficiency and transfer of training?

3. Based upon an examination of the assumptions underlying the safety and job readiness estimates:

   3.1 would any changes contemplated be likely to limit the chances that accident reduction may be achieved?

   3.2 would any changes contemplated be likely to change job readiness estimates?

   3.3 what are the likely cost implications of

changes in the accident reduction estimates or job readiness estimates?

Diagnostic analyses may also aid in designing TD/S to address the more difficult tasks. If a task can be learned (for example, start the engine) in one trial (an easy task) there may be little point in emphasizing it on the TD/S. If it is severable from the learning sequence, i.e., it is not an enabling objective to other tasks, it may be excluded from the TD/S. Similarly, the design may consider criterion reliability and the ability of the TD/S to serve the entire range of students.

Empirical studies of transfer may be reporting results that are biased in the low direction due to failure to differentiate hard vs. easy tasks. Using the TER formula, assume the WS group learns to start the engine and perform procedural tasks on one trial requiring one hour. The TD/S group also learns in one trial of one hour. It is obvious that no time can be saved on the WS. A similar effect would apply on a performance measure. It is unfortunate that more empirical transfer studies have not reported task level transfer data in spite of the risks of unreliability that might be encountered. On GO-NO/GO performance measures, sample sizes are frequently too small to be able to detect differences, but even with larger samples task level analyses go unreported in the empirical literature. Criterion unreliability and student variance also are underreported in the empirical literature. Their masking effects on design features were noted in two studies cited in Chapter 1 by Pfeiffer and his associates (1985).

COST EFFECTIVENESS DECISION RULES IN BRIEF

The Operating Cost Ratio (OCR) presented in detail in Volume II is the basic form of cost analysis of TECIT. It is the life cycle cost per hour of the TD/S divided by the life cycle cost per hour of the WS, or:

$$OCR = \frac{\text{TD/S cost/hr.}}{\text{WS cost/hr.}}$$

When OCR is less than 1.0, the TD/S costs less to operate than the WS. In Orlansky and String's (1979, 1985) reviews of 34 flight simulator studies, the median OCR was .08, showing a very favorable cost ratio. Many TD/S are justified on the basis of a favorable cost ratio and annualized cost savings. Relating costs and effectiveness is, however, not a straightforward matter.

The relationship of the OCR to the Transfer Effectiveness Ratio (TER) is definable because all resource elements necessary for cost analysis of the TD/S and WS are included in the TER formula. In contrast, MAUM-weighted safety ratings, job readiness ratings and performance transfer measures must rely on judgments of (a) whether or not transfer is likely to be achieved and (b) the value of increments of transfer for alternative designs. During the design phase, these questions rely on analytic assessments, while in the fielding phase empirical transfer data may be obtained. When the TER is not an appropriate measure, a general guideline is to design a TD/S to a level of affordability with an OCR less than 1.0, and to maximize expected effectiveness. Iterations of designs and OCRs may then establish an acceptable trade-off point.

Linking the TER and the OCR results in the following decision rules:

1. When TER is equal to or greater than 1.0 and OCR is less than 1.00, the TD/S is cost-effective. Recall that TERs of 1.00 require no additional training time and TERs greater than 1.00 decrease total training time.

2. For TERs greater than 0, but less than 1.00 (the large majority in Orlansky and String's data), the break-even point is when TER = OCR. When TER is greater than OCR, the TD/S is cost effective; when TER is less than OCR, the TD/S is not cost effective. Note that the decision rule cannot be expressed as a cost-effectiveness ratio of equal size units.

3. Cost minimization, assuming performance to criterion is maintained, is achieved when OCR is a minimum relative to TER.

These decision rules are useful in comparing alternative TD/S designs and in task level TER analyses. Alternative designs often consider fidelity design elements which are expected to increase effectiveness but may also be costly to include. Examples are high visual and motion fidelity and computerized response scoring and feedback systems. Analysis of the increments in TER and of costs for the addition of these TD/S elements will yield information helpful for decision making. Task level analyses of TER may be examined for those tasks below the breakeven point and alternatives considered, such as teaching the material in conventional instruction, teaching it on the WS, or improving the TD/S approach to that task.

In contrast to the TER, costs and the Performance Percent Transfer formulae, safety ratings and job readiness ratings follow only a very general set of decision rules:

1. Improve performance without increasing costs.

2. Maintain performance but at a lower cost.

There are, at present, no decision rules or formulae that effectively deal with the situation in which effectiveness increases may be attained but at a higher cost. Whether a given increment in effectiveness is worth an increment in costs is a command decision requiring military judgment.

The reasons for this limitation in associating effectiveness and costs are as follows:

1. The effectiveness measures are often ordinal scales.

2. The value of a particular measure in terms of safety, job performance or battle readiness is not usually established.

On the other hand, the fixed time elements of both the TD/S and WS can be used in relation to the Operating Cost Ratio when performance transfer measures are employed. If there is a relationship between the fixed time required to improve performance on both the TD/S and WS, then these time figures can be used to analyze cost and effectiveness. For example, compare the following combination of times for performance to criterion (or a performance increment) when OCR = 0.2

TD/S: 6 hrs. vs. 8 hrs.
WS: 4 hrs. vs. 6 hrs.

If 8 hours on the TD/S brings the group to criterion on WS in 4 hours, then 8(.2)+4(1.0)=5.6. And if 6 hours on the TD/S requires 6 hours on the WS to reach criterion then 6(.2)+6=7.2. Clearly the first choice (8 and 4 hours) is less expensive. This analysis addresses performance measurement indirectly by analyzing time requirements to improve performance or to reach criterion through the use of a TD/S.

In general, when considering alternative TD/S designs or improving existing TD/S, the aim is to optimize the mix of transfer, performance to criterion, training time and costs. Weighting methods requiring military and technical judgments are needed for this purpose. The Multi-Attribute Utility

Assessment Methods described by Dawdy and Hawley (1985) may be adapted for this purpose.

It should be noted that cost data mature earlier in the life cycle development phases of a TD/S and WS than do effectiveness data. Empirical acquisition and transfer data cannot be obtained until the TD/S is fully fielded, and accident reduction and job readiness follow-up data for some time afterwards, while cost data for the OCR may be reliably estimated by the end of the TD/S development phase. The WS life cycle costs will usually be available earlier. The analyst should bear this in mind in conducting various analyses and in reaching decisions.

TIME, PERFORMANCE, SAFETY, JOB READINESS AND COST TRADE-OFFS

The mix of training time, performance, safety, job readiness and costs are the variables that need to be considered when introducing a TD/S. However, the process is handled slightly differently for time to criterion measures (i.e., TER) and other measures of transfer.

For the TER, the issues are:

1. If TD/S time is truncated because of limits in available total training time, what effect might there be on performance to criterion? The PTC can then be used to measure and assess any deviations from criterion in relation to costs. If no difference in WS time savings (WS-WS(TD/S)) is found or expected, then the truncated TER should be used. As long as the OCR is less than 1.00 and the TER greater than the OCR, the truncated TER will be cost effective. If PTC does make a difference with TD/S truncated, either the required amount of training time will have to be negotiated or performance less than criterion accepted for some percentage of the trainees. The magnitude of the PTC deficit, the amount of TD/S truncation and costs have to be judged jointly.

2. Criterion variability may yield variations above or below the criterion. Using the PTC, estimates of these variations may be made and judged in relation to the TER, PTTS/A and costs.

3. Adjustments in time data for downtime resulting from relative reliability/maintainability may be in order.

4. "Extraneous variance," such as motion sickness in a flight simulator or "one trial learning" for certain tasks on a WS (i.e., the

time for those tasks is not severable from others and there is no "real" WS time saved) can be taken into account in assessing the "worth" of the TD/S.

Recall that TERs greater than 1.00 reduce training time and coupled with OCRs of less than 1.00 are cost-effective. These additional considerations may be most worthwhile assessing when TER is less than about 0.80 or there may be added value beyond that measured by the TER.

When performance measures of transfer, safety, and job readiness are the primary concern, the first issue to be resolved is whether the training when reconfigured by introducing the TD/S would take more or less time and cost more or less than without the TD/S. If training time and/or costs are reduced, the increased performance would add further value to the TD/S. However, this is often not the case. Training time and costs may both increase. Then the issue becomes weighting performance and time increments in relation to costs. Empirical analysis varying time can provide the data necessary for time and performance trade-offs and relative cost, but are not available in the design phase. Military judgment is required to say when the point has been reached when additional performance increments are no longer worth the costs. In the design phase, SMEs can be asked to identify hypothetical trade-off points as an aid to designing the TD/S. Redesign could be indicated if the design does not appear to be achieving the acceptable trade-off bounds.

MULTIPLE COURSE USES AND EXPORTABILITY

While some TD/S are designed for use with a single course of instruction, it is quite common to think of a TD/S as one that has the potential of serving multi-course applications. If devised for "system" training it may be intended to serve more than one course related to that system; and if devised for "non-system" training it may be intended to serve more than one course for a number of different WSs. When multiple applications are envisioned, the TD/S is not a single system. It consists of a core of hardware and software, with courseware and course-specific hardware and software ancillaries available for each course application. Exportable TD/S may often be designed with this flexibility in mind to offer to other branches of the Army the basic hardware and software design upon which couseware can be adapted. In other cases, exportable "packages" may consist of courseware on military skills of sufficient generality that they are expected to have a substantial audience throughout the Army.

It is in this context that the TECIT analytic component can provide additional aid for TD/S evolving through the

development process. Empirical data accumulate at a slow pace as each course application is tested. The TECIT analytic component, on the other hand, can be used to estimate costs and effectiveness for the additional course applications as each application is considered. The structured format will yield much more information (i.e., accident reduction, acquisition, transfer and job readiness estimates) than casual assessments, thus providing estimates specific to the new course application and its environment. This is an application that should not be overlooked.

Chapter 4

RESEARCH STRATEGY AND VALIDATION PLAN

INTRODUCTION

This chapter outlines a number of concepts, assumptions research strategies and a validation plan for tank commander armor training for the TECIT training effectiveness submodel. As noted in Chapter 1, the TECIT model has been developed for use at all phases of the TD/S development life cycle, uses both analytic and empirical methods, and provides a means for joint consideration of applications and research. Accumulation of TECIT analysis may then form a useful database of combined analytic and empirical methods useful for improving the TD/S development process.

It is not intended that applications of TECIT wait until all the research evidence is in. Validity is accumulated incrementally. Many training, educational and psychological models accumulate research data in tandem with their application while some models accumulate application data as one method of research. Field validation studies require timely and appropriate field opportunities and cooperation in operational settings. Thus, the model, documentation of applications, and validation are expected to evolve in conjunction with one another. As experience is gained with the model and validation research accumulates, TECIT will be improved. The documentation of the basis for design decisions, forecasts and validation studies will become part of an accumulating database.

The central research issues are: (1) What is the validity of analytic estimates using TECIT methods? (2) What methods and aids can be employed by analysts to make them more accurate? (3) To what extent, under what circumstances, and for what applications are analytic estimates a useful complement to empirical data? (4) To what extent and for what applications can analytic estimates serve as a proxy for empirical data?

CONCEPTS AND ASSUMPTIONS

The following concepts and assumptions are important to an articulation of research strategies and methods.

1. Model applications differ at various TD/S life cycle phases. Applications differ largely in regard to the conceptual, design and development phases vs. the fielding phase. In the early phases of TD/S development the applications are concerned with the following: Is a TD/S needed? What knowledges and skills can be taught most cost-effectively on the TD/S vs. conventional training? Which of two (or more) TD/S concepts or designs are likely

to be the most cost-effective? These questions help to formulate a set of specifications for a statement of work for bid by contractors, to evaluate competing proposals and to select a contractor to develop the TD/S.

In the development phase, as the design begins to evolve, analysts and contractors may use the model to aid in making decisions related to the cost-effectiveness of development alternatives. As the fielding phase approaches, planning, installation, deployment and empirical studies become paramount concerns. The early phases operate without data on the specific TD/S under development while the fielding phase begins the accumulation of empirical data.

2. Risk and uncertainty lead to reserves for contingencies that vary according to the TD/S life cycle phases and baseline availability of information appropriate to various applications. In the conceptual and design phases of TD/S development, by definition, there is no empirical information available about the TD/S and uncertainty is high. However, most of the major design decisions are made.

In the fielding phases, the design is largely fixed. Although empirical data begin to accumulate, they do so at a slow pace. There are still many areas of risk and uncertainty regarding the installation, deployment, utilization and effectiveness of the TD/S for various courses and applications.

Because of these risks and uncertainties there is a tendency to think in terms of reserving judgment and resources for contingencies. These contingencies may be related to factors external to the TD/S or internal to the TD/S. External factors may include changes in the threat scenario, the WS(s), the training programs, policy or doctrine. Factors internal to the TD/S may include those resulting from lack of information at the concept and development phases and the changing state of the art such as those provided by emerging computer-based technologies. It is hypothesized that methods for reducing risks and uncertainties should result in a concomitant reduction in reserves.

3. In general, valid and reliable information aids in reducing risk and uncertainty and should reduce concomitant reserves. Formal models such as TECIT, DEFT, FORTE, and CBP are designed to aid in reducing risks, however, how effectively and the extent to which they do so has not been thoroughly researched and is the subject of this chapter. Models and methods for reducing risk may be oriented only to factors internal to TD/S development (i.e., TECIT, DEFT, FORTE), may take WS(s) and training program development into account (i.e., Training Effectiveness, Cost Effectiveness Prediction, Training Developers Decision Support System) or

include consideration of WS development, manpower, personnel and training in the conceptual phase of WS development.

Baseline methods such as the use of databases, meta-analyses, predecessor and similar TD/S when available, appropriate, and properly interpreted may be used to reduce uncertainty. Systems analytic methods, task analyses, sensitivity analysis, expert judgment and statistical estimating procedures may also aid in this regard. The research issues are: (a) how should the methods be used and combined most productively and (b) how valid is each method for various applications. The availability, cost and value of these sources have not been adequately explored.

4. The criteria against which model estimates are validated are the empirical measures obtained after the TD/S has been fielded. These data mature in about the following sequential order.

a. Acquisition learning (validation/verification, pilot study) empirical study.

b. Reliability of student performance on the TD/S and the WS exercise.

c. Reliability/maintainability of the TD/S.

d. In-course transfer of training study to WS exercise.

e. Utilization rates of the TD/S.

f. Skill decay and skill maintenance analysis.

The empirical measures are fallible (i.e., contain their own error variance) and thus represent partial criteria.

5. Model metrics and analysis methods should lend themselves to validation methods. Since TECIT enables reliability and variance estimates to be made with respect to time and performance in acquisition and transfer, its metrics readily lend themselves to validation when a TD/S is first fielded. These in-course validation studies should lend partial credence to the efficacy of the model. Follow-up and long-term studies are needed to validate the model for job readiness, safety and the utilization ratio.

6. Model validity methods need to focus on predictive validity and accuracy. In traditional psychological and educational measurement theory, face validity refers to the extent to which an instrument appears to be measuring what it is supposed to measure. In other words, do the items in the test or questionnaire appear to be measuring directly or indirectly what the author claims they measure? For a cost and training effectiveness analysis model, face validity is a bit more stringent. Face validity refers to the reasonableness of all elements of the model taken separately

and together. In the case of TECIT, face validity includes the formulation of the applications appropriate to various life cycle phases, the training spectrum analysis, and other aspects of the Problem Definition and Analysis Component. Face validity also includes the effectiveness function and its definition, the metrics employed, the judgmental variance concept and all other aspects of the model taken together. Review of the initial model by experts in TD/S modeling and development for clarity of definition and procedure may support the model to varying degrees, suggest limitations, and suggest means for improving it. Nonetheless, most if not all CTEA models (at least those reviewed recently) easily pass the test of face validity.

Operational validity refers to the apparent usefulness of the model. To the uninitiated user, the appeal of a model may lie in its apparent utility, ease of use and the perceived value of the models output in aiding decision making. If the definitions and methods are sufficiently clear, the model will be operationally useable without excessive difficulty and yield information of apparent value. Since TECIT is a multi-purpose and multi-application model, operational validity can accumulate only as experience is gathered in its application to the variety of problems for which it was designed.

While face and operational validity are important first steps in establishing the validity of a model, only empirical validity methods demonstrate a model's ability to predict or to discriminate in measurable ways.

Empirical validity methods relate judgments to empirical data or to known characteristics of a TD/S. Methods include predictive validity, concurrent validity, discriminant validity and convergent validity. Predictive and discriminant validity are the most important at all phases of TD/S development. Statistical methods appropriate to their measurement include correlation and comparison of averages.

Accuracy of analytic estimates of time and performance as opposed to correlations of analytic and empirical data is the more stringent validity measure. Correlations show only whether analytic and empirical measures tend to follow the same rank order. While correlations are a useful measure of validity, they do not show the degree of accuracy of analytic estimates of training time and performance. Training time inaccuracies affect both instructional management and cost estimates. Much of cost estimating depends on the time over which resources are used.

The practical consequences of overestimating vs. underestimating time and performance differ a great deal. Overestimating provides resource reserves for contingencies, while underestimating may result in ineffective performance and inadequate resources. The research on FORTE discussed

in Chapter 1 amply illustrates the distinction between accuracy and correlation.

7. Sources of error variance of analytic methods need to be articulated and analyzed systematically to develop a framework for testing hypotheses. By identifying analytic error variance sources, methods may be directed to controlling them or taking them into account when employing analytic methods. Some example of hypotheses regarding analytic error variance sources are as follows: Error variance (i.e., the discrepancy between analytic and empirical time and performance estimates) is expected to be greater when: (a) estimates are made while the WS and training program are still in development; (b) there is little information or inconsistent information available related to the TD/S design; (c) analysts and SMEs are inexperienced; (d) different analysts and SMEs are used at various phases of TD/S development; and (e) the "state-of-the-art" (i.e., computer technology) in TD/S design is changing.

It should be noted that concurrent validity studies may minimize or control sources of error variance such as those associated with changes in the WS and training program and are useful for this purpose. However, they also tend to minimize wanted variance. For example, one may want to vary information input and SME qualifications and analyze the effect on the predictive accuracy of analytic estimates over time. Concurrent validity studies confound the time variable.

8. Empirical data are fallible. While empirical data are important as criteria for analytic studies and more rather than fewer empirical studies are needed, empirical studies themselves may be limited by small sample sizes, lack of replication, confounding of treatments, and biased data resulting from inappropriate measures or other threats to validity. Inferences from a transfer experiment to the population of users may or may not represent true differences between a transfer and control group.

9. Content validity (i.e., the content is judged valid by experts) of a TD/S is particularly important when safety and battle readiness are key areas for which the TD/S is designed. The reasons for this are as follows: (a) Criteria for safety and battle readiness do not mature for many years after a TD/S is fielded. Short-term measures may be misleading. (b) Attributing safety and battle readiness to a particular TD/S against the backdrop of other training and experience is difficult to do with the available "state-of-the-art" study designs.

Emphasis has been on analytic methods and military judgment related to incorporating critical safety or battle simulation content in a TD/S. Some examples include: (a) reacting to simulated wind shear situations in flight

training; and (b) increasing chances of survivability, hits and kills in tank training.

Adaptive TD/S have the capability of modifying software and courseware to accommodate newly recognized safety problems and battle readiness scenarios without necessarily changing the hardware configuration. This capability by definition enhances content validity of the TD/S courseware. There are at the same time effectiveness and cost implications which have not been explored. How much is effectiveness expected to increase? What are the development and operational costs associated with hardware and software flexibility and with changes in the courseware to accommodate newly perceived threats? These areas are worthy of further research.

10. Samples of study team members, analysts and SMEs of sufficient size are needed to provide tests of reliability, validity, and accuracy of prediction. As opportunities for application of the model arise, researchers should make every effort to assure that the size of the study team and analysts is large enough to be able to compare important variances. Primary attention should be given to study team members and analysts as they are responsible for structuring the analysis and making the analytic estimates.

The analysts' task is to define and analyze the problem and to identify the need for SMEs where appropriate. SME sampling is important in areas in which it is unlikely that team members or the analyst will have expertise, or to cross-check analyst estimates. As the configuration of areas of experience and expertise is quite large and practical samples generally quite small (2 to 30), the researcher should develop a clear notion of the most important analyses and comparisons to be made.

11. Evaluability Assessment. The effectiveness submodel of TECIT makes one very important assumption central to its use: that criterion measures of time and performance on the WS or job can be made of sufficient reliability and validity to enable forecasts to predict them accurately. Early attention to the reliability and validity of the criterion measures on the WS will tend to obviate differing forecasts by clarifying the intended outcome measures. Without this clarity of measurable outcomes, the effectiveness of TD/S designs will remain ambiguous and unevaluable by quantitative methods, with perhaps commensurate tendencies toward overdesign, high costs, and unknown relationships to military readiness.

RESEARCH STRATEGIES

In general, cross-sectional, baseline, longitudinal, and joint analytic and empirical strategies are appropriate to research on analytic models.

Figure 4 shows a general structure for the cross-sectional, baseline, and longitudinal research strategies. This structure shows the three strategies in relation to the applications of various TD/S life cycle phases (concept and design vs. fielding) and in relation to elements of the TD/S effectiveness function.

Cross-sectional validity strategies. These strategies are most useful in relation to the concept and design phases of TD/S development. Analyzing the model in relation to a sample of TD/S of known characteristics provides opportunities to determine how well the model can discriminate between the characteristics of various TD/S (i.e., those that are part of the TD/S effectiveness function plus characteristics of the training program, physical and functional fidelity and instructional management).

This strategy requires calibrating information of sources of variances from different existing TD/Ss to aid judgments in other TD/S designs. A sample of TD/Ss is selected with known characteristics such as various transfer results, TD/S designed for safety, transfer, criterion referenced TD/S for battle conditions, utilization ratios, etc. Hypotheses are developed about each TD/S in the sample. The TECIT measures for common hypotheses are then devised, including variances such as student ability, instructor leniency, task difficulty, physical and functional fidelity, team variables, and other identifiable sources of variances. SMEs who are not familiar with the empirical data and hypotheses make estimates for each TD/S and set of variables. These results are then analyzed empirically in relation to the hypotheses. The relationships will yield a better understanding of what SMEs are capable of forecasting and calibration of important variables.

At this juncture, there does not appear to be a straight forward and simple method for establishing minimum standards based on concept and design characteristics. The number and range of design variables, the complexity of their interrelationships and the lack of empirical knowledge relating design characteristics to effectiveness measures make the task of specifying minimum standards very formidable. Nonetheless, it is a subject which requires further research attention and would be an important aid in the formulation of TD/S concepts and designs.

Baseline Strategies. While cross-sectional studies are useful, taken alone they may be sterile as they may demonstrate the validity of the model but fail to provide benchmarks to analysts to improve the basis for their estimates. Development of baseline methods for analysis in conjunction with the TD/S sample would be most productive in providing benchmarks for consideration by analysts in the

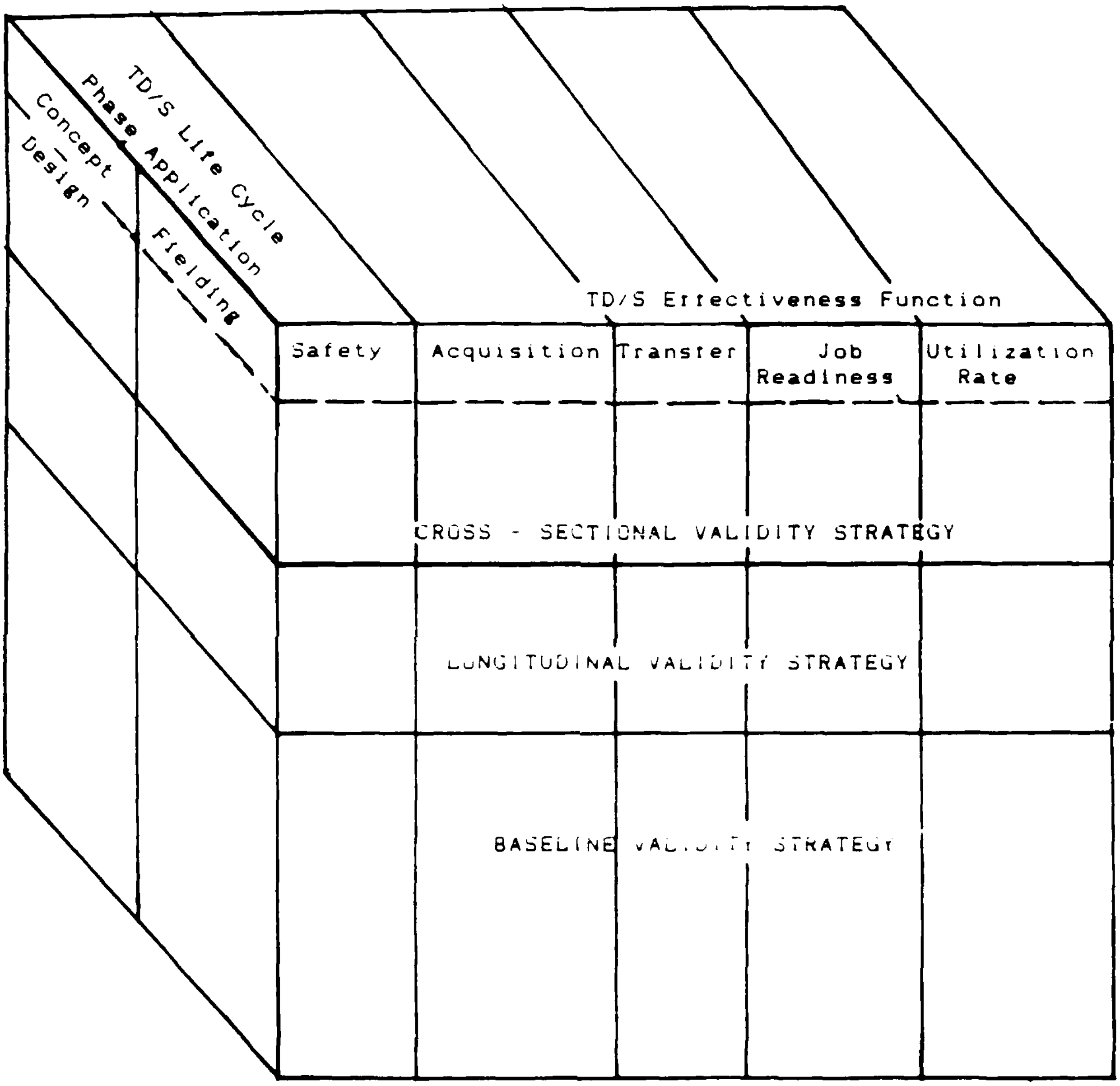

Figure 4: Research Strategies, TD/S Life Cycle Phase Applications and Elements of the TD/S Effectiveness Function.

concept and design phase. Databases, meta-analyses and comparison-based methods are useful in this regard.

Databases containing summaries of empirical studies, when available, will be a useful part of an analytic model. Using databases, estimates might be refined and SMEs judgments cross-validated. At present, only limited data bases exist for use with TD/Ss. Orlansky and String's (1977, 1979, 1985) studies of the cost-effectiveness of flight simulators, maintenance simulators and computer-assisted instruction are illustrations. The results of their effectivness measures are presented in Appendix A. It is our understanding that these data are being updated and that other databases are being prepared in the Army and throughout DOD.

The Orlansky and String database is useful in examining summary results, particularly Transfer Effectiveness Ratios (TERs) and Percent Time Saved (PTSs) for flight simulators. The similarity to planned TD/Ss and the range of values may help guide estimates for new TD/S. However, summary data of this type are somewhat limited for designing TD/Ss as they do not give insight into the design features that contribute to marginal changes in effectiveness or costs. Comparison-based methods may be useful in this case.

The type of databases that are needed would yield expected values of training effectiveness as a function of task difficulty, physical fidelity features (i.e., visual, motion, etc.) instructional features and student characteristics. While a comprehensive database is a long way off, improvement of the available information could be made by means of literature reviews, programmatic research and indexing studies of existing TD/S in relation to the variables of interest. For example, Rose and Wheaton (1985) in their literature review leading to the development of the Device Effectiveness Forecasting Technique point out the importance of task difficulty and the number of "steps", physical and mental, in learning a task. Blaiwes and Regan (1986) report that Evans and his colleagues at the Naval Training Devices Center are conducting research to test the effects of various fidelity features. Studies of this sort should provide useful leads in designing TD/S.

The generalizability of databases will always be open to question when attempting to translate findings from one WS or job to another (for example, flight simulators to tanks), from old to new technologies (for example, mechanical to computerized sub-systems) or from old to new instructional concepts (for example, the renewed emphasis on cognitive processes as mediators to performance generalization and transfer). However existing databases provide researchers with useful guidance in defining the requirements for new databases and empirical studies. Further, they may provide analysts with preliminary data that can aid in establishing the cost and effectiveness bounds of a proposed TD/S design.

Longitudinal Strategies. Short-term (two to four years) longitudinal strategies are another viable alternative. For example, a longitudinal strategy during the concept, design and development phases may be useful in tracing the evolution of a TD/S. Comparisons of first and last designs as a function of information input would be of interest in testing the cost and value of information.

Long-term longitudinal studies are of limited value. The lead time (often three to ten years) and resource requirements to carry out such studies from concept through fielding make such studies untimely. Changes in the threat scenario, policy and the WS are difficult to control.

Long-term longitudinal validation strategies also confound iterations of the model, changes in the WS, changes in the training program and changes in the TD/S concept.

On the other hand, when a TD/S, WS and TP are all fielded, the probability of changes is greatly reduced (though not zero) providing control for these variances. The lead time (six months to four years) and resources required for a longitudinal study become much more practical and can yield insights into the bases of SMEs forecasts of importance in testing the model.

In the fielding phase, short-term longitudinal validation designs include predictive and concurrent validity studies. The last time and performance estimates (acquisition and transfer) of the development phase may be used as predictors of empirical time and performance acquisition and transfer measures. Further follow-up during and after the course can be undertaken in relation to safety, job readiness and the utilization ratio.

Concurrent or follow-back studies are of some interest in controlling variances, but it should be kept in mind that they may represent judgments after important decisions have been made. There is the attendant risk that the SMEs judgments may not be independent of one another if "the word has gotten out" about certain decisions or if a bias develops outside of the study setting. The risks of SME non-independence and bias are most obvious when a group of instructors from the same school or researchers from the same unit serve as SMEs. Independence of raters has to be judged in balance with familiarity with training issues related to the TD/S, TP and WS. There do appear to be ways of balancing these concerns and detecting contaminating non-independence and bias. For example, effects of group differences such as instructors, researchers, psychologists, engineers, etc., can be tested in relation to varying information inputs and the extent to which groups have worked together in the past. Studies of this type might be done using university students, faculty, public school

vocational teachers and contractor "experts" as well as SMEs who work for the military.

The post-development longitudinal validation strategy is limited in one important respect. It will be applied to situations in which many of the major design decisions have already been made, limiting its value to a restricted range of variances. It is likely to be less useful for decisions related to TP vs. TD/S tradeoffs, design decisions related to marginal changes in TD/S fidelity and costs, and screening tasks for cost effectiveness. In fact, the more valid are early design decisions, the more restricted will be the alternatives that can be assessed at post-development phases of the TD/S life cycle. Thus, the range of discriminability of alternatives that distinguish a profile of acceptable vs. unacceptable designs cannot be fully addressed.

Joint use of empirical and analytic data. In the fielding phases when empirical data become available, these data may be limited as noted above. For certain purposes, studies should be undertaken to determine the extent to which analytic data can complement empirical data and the extent to which analytic data may serve as a proxy for empirical data. FORTE studies discussed in Chapter 1 suggest this may be possible. A number of applications are suggested as a starting point as follows:

(a) Small sample sizes in the transfer experiment. When sample sizes are small, it is difficult to discriminate statistically between true differences in the transfer and control group. Increasing sample size may not be an operationally affordable solution. A research approach would provide SMEs with the transfer data and ask them to extrapolate the results to the larger student population, asking whether they would expect the results to be about the same, higher or lower than the data obtained. A later replication of the transfer experiment would then be compared with the analytic data. Note that this approach differs from a forecasting approach. This type of study may be conducted when sample sizes are expected to be sufficiently large, but trainees become available in various training cycles.

(b) Confounding of treatments in a transfer experiment. In many cases, it is not operationally possible to conduct a transfer experiment in which all possible treatment combinations are included that one would like. A research approach would first compare a confounded treatment group with a control group. For example, the Pfeiffer and Associates (1985) study discussed in Chapter 1 might have used the visual plus motion fidelity group vs. the no-visual-no motion group. Second, SMEs would then be presented with the resulting empirical data and asked to interpolate the results they would expect from each treatment separately. Third, the unconfounded treatments

would then be tested empirically and compared with the analytic results.

(c) Extrapolation for exportable packages and multi-course uses. Exportability and multi-course applications suggest that empirical results for a TD/S for one course and one setting will apply to another course and another setting. Empirical studies thus take on the tone of demonstrations of the potential value of a TD/S. This inference is quite natural, however, it would be useful to formalize the extrapolation by taking account of variances that may differ in the new settings and courses. If transfer occurs in one setting, it may be higher or lower in another.

VALIDATION PLAN

Background

The development of two new TD/S and a new exercise on the WS at the Ft. Knox Armor School provides an opportunity to validate selected aspects of the TECIT Model for the Tank Commander's (TC) Basic Non-Commissioned Officer's Course (BNCOC) for the M1 Abrams tank. They are Simulated Combined Arms Training (SIMCAT) and Computer-Assisted Instruction (CAI) lessons.

The applications of TECIT proposed include: joint analytic and empirical studies of acquisition learning, instructional management, transfer of training within the TC course, estimation of battle readiness, exportability analysis and life cycle cost analysis. These applications are for a WS that has been fielded and TD/S that are in advanced phases of development. The M1 Abrams tank and the TC BNCOC program of instruction have been fielded for many years. The two TD/S, the CAI lessons and SIMCAT, and the new WS exercise are now nearing completion by contractors. As of this writing, they were expected to be ready for full delivery by about late fall or winter 1986.

Description, Purposes and Expectations of TD/S and WS Exercise

SIMCAT is a generic simulator focusing on command, control and communication skills. Off-the-shelf hardware is being used. Courseware is being developed for the TC BNCOC and several officers' courses. A brief description of the SIMCAT exercise for the TC course is presented in Table 17. The courseware provides a "free-play" exercise of a simulated battle environment which includes friendly and opposing forces in various configurations. The SIMCAT exercise fits the characteristics of a work sample simulator discussed in Chapters 1 and 3. It samples from a variety of battle conditions which may be infrequently encountered and may not otherwise be readily presented in training. Thus,

Table 17

Simulation in Combined Arms Training (SIMCAT)

1. SIMCAT is a computer-based platoon-level battle simulation developed by the Army Research Institute (ARI) to support armor training research. There are plans to use SIMCAT to produce effective and efficient methods for training command, control, and communcation (C) skills and platoon-level tactics.

2. SIMCAT allows up to four participants to serve as TCs (Tank Commanders) of simulated M1 tanks. Each TC has a computer monitor display which indicates the location of his tank and any other vehicles which would be in line of sight. The location and orientation of each vehicle is indicated by a computer-generated graphic icon which is superimposed at the appropriate location on a map display.

3. Each SIMCAT TC station contains a microcomputer which can recognize human speech. The TC issues voice commands to control the movement and firing of his tank. For example, the TC can say "Driver, MOVE OUT" and his vehicle will begin to move on the display screen. The actions of the gunner, driver, and loader are simulated by computer.

4. Platoon and Company Communication nets allow practice of standard CEOI procedures. For communication purposes a Chief Controller serves as the Company Commander. This controller also represents the FIST during calls and adjustments for indirect fire.

5. An OPFOR controller commands T72s, and BMPs with SAGGERS, to provide an active, intelligent threat. The OPFOR controller can also employ indirect fire and can place minefields at any point on the 5 x 20 kilomether battlefield.

Source: ARI Field Unit at the Fort Knox, Kentucky Armor School, Courtesy D. M. Kristiansen

the variables of interest in validating its effectiveness include both battle readiness estimates and transfer of training within the course. In-course transfer depends on the appropriateness of the WS exercise to the SIMCAT exercise. Thus, the SIMCAT exercise and the WS STTX may have common variance (i.e., transfer within the course) as well as unique variance with regard to improved preparation of trainees for battle.

At this writing, performance measures, the GO/NO-GO criterion, training time requirements, and course scheduling are still being considered. A preliminary training time estimate for the TC course given by the project officer is eight hours. An empirical study is required to determine the effect of time variation on performance on the STTX. Instructional management issues are still unsettled providing opportunity for validation of that scale.

A life cycle cost analysis is needed to estimate the Operating Costs/Hour (OCH) for SIMCAT for comparison with the costs on the WS (see Volume II).

The CAI lessons in preparation include a group of lessons of common military skills (i.e., Communication Electronics Operating Instructions, Land Navigation, Land Navigation Using Surrogate Travel, and NBC Warfare) and other lessons more specific to the TC course (i.e., Remediation, Mine Warefare, and Call For/Adjust Indirect Fire). A detailed outline is given in Table 18. The common military skills lessons are intended to be exportable packages of instruction potentially useful in many other Army training settings. Hence, validation in the TC course will provide a benchmark for extrapolation to other training environments. The validation of TC-specific lessons will demonstrate the validity of the CAI approach in improving performance in that course but will not necessarily be useful in other courses.

The CAI lessons are viewed as part-task training and are expected to demonstrate transfer of training to the exercise on the M1 Abrams tank. It is of interest to note that a transfer study using CAI lessons would be the first of its kind. Orlansky and String's (1977, 1979, 1985) reviews of CAI effectiveness studies found that all studies reported compared CAI lessons with conventional classroom instruction. None of the studies examined transfer of training. Very few studies examined the cost effectiveness of CAI, and then with limited cost models.

The CAI lessons are presented on the Micro-TICCIT system coupled with the Videodisc. A unique feature of the couseware is its emphasis on graphics, motion and audio presentation. This approach to courseware is expected to be motivating to the trainees, improve learning, and avoid reliance on higher level reading and verbal skills.

Table 18

Computer-Assisted Instructional Units by Task Cluster and Sub-Task

Communications Electronics Operating Instructions	
	Item Identifiers
	Call Signs
	Suffixes
	Frequencies
	Encoding
	Decoding
	Authentication
	Radio Procedures
Land Navigation	
	Determine Grid Coordinates
	Analyze Terrain Using Five Aspects of Terrain
	Identify Natural Terrain Features
	Determine Elevation
	Orient Map to Ground by Terrain Association
	Determine Location by Terrain Association
	Locate an Unknown Point by Intersection and Resection
Land Navigation Using Surrogate Travel	
	Determine Location by Terrain Association
	Navigate from One Point on the Ground to Another
	Reconnaissance by Surrogate Travel
Fire Commands	
	Stationary Tank, Stationary Target
	Stationary Tank, Moving Target
	Stationary Tank, Multiple Targets
	Simultaneous Engagements
Remediation	
	Determine Grid Coordinates
	Communicate Using Visual Signaling Techniques
	Recognize and Identify Friendly and Threat Vehicles
	Establish Tank Firing Positions
Nuclear, Biological, and Chemical Warfare	
	NBC Reporting
	Radiacmeter
	Dosimeter
	Chemical Kit
Mine Warefare	
	Install a Hasty Protective Minefield
	Direct a Minefield Marking Party
Call For/Adjust Indirect Fire	
	Range Estimation
	"Mil" Formula
	Grid Missions
	Shift from a Known Point
	Polar Plot

Source: ARI Field Unit at the Fort Knox, Kentucky Armor School

A number of CAI lessons have been delivered and are undergoing preliminary validation with a small sample of trainees. The preliminary validation is comparing performance (pre-test and post-test) and learning time for trainees completing CAI units vs. trainees who participate only in conventional classroom instruction. This preliminary validation addresses acquisition learning but does not address transfer of training. Instructional management is being given attention, but there are as yet issues of scheduling that have not been resolved.

A new field training exercise (STTX) is also being developed for the TC BNCOC course. An outline of this exercise is presented in Table 19. This STTX is also expected to be available about late 1986 or early 1987. The purpose of this revised exercise is to provide more realistic battle training than the one it replaces. According to the project officer it is expected to require about 1.5 hours of instruction per TC compared with 4.5 hours of instruction for the old exercise. Although the content has been outlined, performance measures and the GO/NO-GO criterion have yet to be established and tested for reliability. The new STTX is also intended to serve as the criterion measure of in-course transfer of training for SIMCAT and the CAI lessons.

As noted in the discussion of SIMCAT, the new STTX is expected to have unique variance in improving battle readiness as well as variance in common (i.e., transfer) with SIMCAT.

Study Design

A series of five studies is recommended as follows:

1. Predictive validity of analytic to empirical acquisition learning for CAI lessons, the SIMCAT Tank Commander exercise, and the STTX.

2. Joint analytic and empirical study of in-course transfer of training from CAI and SIMCAT to the STTX on the M1 Abrams tank. Concurrent validity and interpolation of empirical data will be obtained for analytic estimates.

3. Follow-up validation of the Utilization Ratio scale.

4. Predictive validity of the battle readiness measures.

5. Cost and cost effectiveness analyses

The five studies are expected to require three years to complete because of the lag time involved for empirical data to mature. Studies 1, 2 and 5 can be accomplished within 1-1/2 years, and studies 3 and 4 within 3 years. In

Table 19

STTX Task by Station for the Tank Commander Basic Non-Commissioned Officers Course on the M1 Abrams Tank

STA#	TASK
1	PREPARE FOR OPERATIONS
	GS-10. Install/remove .50 cal. machine gun; GS-20. Prepare CWS for operation; GS-23. Perform commander's prepare-to-fire PMCS: GS-12. Boresight .50 cal. machinegun; GS-11. Zero .50 cal. machinegun; GS-31. Boresight/system calibrate M1 tank; LN-11. Identify adjoining map sheets; T-5. Conduct troop leading procedures; T-7. Prepare and issue an oral operation order; C-5. Use an automated CEOI; C-1. Establish, enter, leave radio net; C-3 Use KTC 1400 numerical cipher/authentication system.(12 tasks)
2	ENGAGE OPFOR TANK FROM CWS
	GS-22. Engage targets w/main gun from CWS; LN-5. Orient a map to the ground by map/terrain association; LN-2. Determine location on the ground by terrain association; C-2. Encode/decode message using the KTC 600 Tactical Operations Code.(4 tasks)
3	OPFOR INDIRECT FIRE
	NBC-6. Implement MOPP (2 to 4); NCE-2. Use M256 chemical detector kit; NBC-3. Initiate unmasking procedures; NBC-6. Implement MOPP (4 to 2). (4 tasks)
4	REPORT OF NUCLEAR BURST
	LN-1. Determine magnetic azimuth using a compass; NBC-5. Prepare/submit NBC-1 report; NBC-4. Use IM-174 radio nets; C-2. Encode/decode messages using KTC 600 Tactical Operations Code; NBC-7. Prepaare/submit NBC-4 report. (5 tasks)
5	OPFOR MACHINEGUN FIRE AT ROAD OBSTACLE
	GS-8. Engage targets with the M240 coax from CWS. (1 task)
6	OPFOR TANK BLOCKING ROUTE OF MARCH
	T-3. Call for and adjust indirect fire; GS-25. Direct main gun engagements on an M1 tank. (2 tasks)

Table 19 (cont'd.)

7	POSSIBLE CHEMICAL CONTAMINATION

NBC-6. Implement MOPP (2 to 4); NBC-2. Use M256 chemical detector kit; NBC-3. Initiate unmasking procedures; NBC-6. Implement MOPP (4 to 2). (4 tasks)

8	OPFOR ELECTRONIC COUNTER-MEASURES

C-4. Recognize electronic counter-measures (ECM) and implement electronic counter-counter-measures (ECCM). (1 task)

9	OPFOR SNIPER AT ROAD OBSTACLE

GS-9. Engage targets with .50 cal machinegun; LN-8. Determine azimuth using a protractor and compute back azimuth; LN-7. Locate an unknown point on a map or on the ground by resection. (3 tasks)

10	OPFOR TANK ENCOUNTER DURING ROAD OBSTACLE BYPASS

LN-5. Orient a map to the ground by map/terrain association; GS-25. Direct main gun engagements on an M1 tank; C-2. Encode/decode messages using the KTC 600 Tactical Code. (3 tasks)

11	DEFEND BATTLE POSITION AND CLOSE OPERATIONS

T-2. Select a firing position; NBC-1. Read/report radiation dosages; LN-4. Orient a map using a compass; LN-10. Identify terrain features on a map; LN-9. Analyze terrain using the five military aspects of terrain; T-1. Install/remove hasty minefield; GS-24. Direct machinegun engagements; T-4. Estimate range; LN-6. Locate an unknown point on a map or on the ground by intersection; T-3. Call for and adjust indirect fire; GS-25. Direct main gun engagements on an M1 tank; T-1. Install/remove hasty minefield; C-3. Use KTC 1400 numerical/cipher authentication system; C-1. Establish, enter, leave radio net; GS-21. Secure CWS; GS-10. Install/remove .50 cal. machinegun; LN-3. Navigate from one point on the ground to another point. (17 tasks)

12	NIGHT OPERATIONS

Occupy night defensive position, move into hull defilade position; Report platoon sector OPFOR activity, OPFOR tank engine startup, exposed OPFOR TANK, OPFOR flare, OPFOR firing machinegun

Source: ARI Field Unit as the Ft. Knox, Kentucky Armor School. Tasks are in sequence within station

general, the analyses will tell whether analytic estimates can substitute for long term data gathering.

Study 1. Predictive and concurrent validity of analytic to empirical acquisition learning of CAI lessons and the STTX.

As noted earlier, the TD/S and STTX are in advanced stages of development and will be undergoing empirical validation testing on small samples, to confirm time requirement estimates and establish performance measures and criteria. Selected CAI lessons may also be compared with conventional classroom instruction.

After performance criteria have been specified, the predictive validity study of analytic estimates will ask SMEs to predict performance for a larger sample, all trainees over a one-year period, based on the small sample (5-15) results. Empirical data will be accumulated for the one-year period to serve as the empirical criterion measure. The analysis will obtain judgmental variances for CAI lessons and STTX stations and selected tasks; student variances for the STTX are as follows:

Table 20

Judgmental Variance Sources for Acquisition Learning

CAI	Lessons	Students	Time to GO; Percent by-passing each lesson
STTX	Stations, Selected Tasks	Students	Instructor Variance

SIMCAT is not included as it is a free play exercise that reportedly will vary a great deal in the experiences encountered and the judgmental scoring of the exercise.

It is recommended that the analytic estimates be made by two groups of personnel to enable comparisons to be made of ability to predict and to conduct tests of inter-rater reliability. The two groups and recommended samples are: (1) ARI, TTFA and contract developers - two to five personnel including those ARI/TTFA/contractor personnel familiar with the development of each item; (2) Tank Commander BNCOC instructors - two-five involved in the installation and validation of the CAI and STTX.

Comparison of results will be as follows:

1. CAI - for each lesson and overall

1.1 Mean and standard error of time to "GO" from small sample, full-year sample and analytic estimate. The correlation and accuracy of the mean and standard error of the analytic estimate and the full year empirical data will provide the test of predictive validity. These analytic-to-full-year comparisons will also be compared to small sample-to-full-year data to test the efficacy of the analytic estimates.

1.2 Percent by-passing each lesson. As before, comparisons and correlations will be made of analytic and full-year empirical data with small sample to full-year correlations and accuracy estimates.

2. STTX - for each station, task and over-all

Obtain the mean and standard error of "GO's" for the overall score for the small sample, the analytic estimates and full-year data. Analyze station, student and instructor variance in relation to full-year data for predictive validity and accuracy. Compare analytic predictions to small sample predictions.

Analytic exportability estimates may also be obtained for those CAI lessons developed for common military skills. However, validation of analytic estimates may not be possible for lack of empirical data.

<u>Study 2</u> - Joint analytic and empirical study of in-course transfer of training from CAI and SIMCAT to the STTX on the M1 Abrams tank.

It is well known that acquisition learning on a TD/S does not by itself establish effectiveness. It demonstrates only that trainees learn on the TD/S. Transfer of training to an exercise on the WS is the more convincing demonstration of the effectiveness of a TD/S, particularly when there is an appropriate exercise within the training program against which to measure transfer. Analytic estimates of transfer would be useful and less costly if it can be demonstrated that the analytic estimates accurately forecast transfer.

Empirical transfer studies pose certain difficulties. In the present instance, the number of trainees enrolled in the TC BNCOC course is quite small, typically 12-16 per class with an annual throughput of 96 to 128. This small population size poses difficulties in obtaining empirical transfer data particularly when there is more than one treatment group. In this instance, one would want at least four treatment groups as follows:

1. control group
2. CAI only group
3. SIMCAT only group
4. both SIMCAT and CAI

Additional subgroups can be identified depending on the order in which CAI and SIMCAT are presented.

Given these circumstances, a lagged empirical transfer design with analytic interpolation for lagged treatments and extrapolation for full-year data provides a practical means of obtaining empirical transfer data and validating analytic estimates.

The study would proceed as outlined in Table 21. Note that the confounded treatment in number 1 has the advantages of being the most timely, yielding short-term results and enabling experience to be gained with all methods. The control group is not deprived of CAI and SIMCAT; they simply take them in a different order. However, it has one major disadvantage, namely separate estimates of transfer effectiveness are not obtained to relate to costs. This issue is resolved in two and three and the validity of analytic estimates obtained.

Four classes will be required for the empirical studies. Considering class scheduling, the data gathering may be accomplished in six to seven months.

The dependent variable is the performance measure on the STTX for each station, task (or subgroup) and overall. Time on CAI and student characteristics may be used as covariates in an Analysis of Covariance Design. Performance transfer formulae may be selected from among those discussed in Chapters 1 and 3, namely the Percent Transfer to Criterion (PTC), Percent Transfer to Maximum (PTM) or others.

<u>Study 3.</u> Follow-up validation of the Utilization Ratio Scale.

This study will obtain two SME estimates of the Utilization Ratio scale presented in Chapter 3. The first estimate will be made in year 1, the second in year 2. Estimates are to be made by ARI Field Unit and TTFA personnel. The estimates would be correlated and compared with actual utilization ratios gathered over 2 1/2 years.

<u>Study 4.</u> Predictive validity of the battle readiness measures for SIMCAT and the STTX.

As noted throughout Chapters 1 and 3, many TD/S and WS exercises are work samples of realistic battle conditions that may be expected to improve transfer to the job after training, but whose effectiveness may be measured only in part by a transfer study while in the training program.

Table 21

Joint Empirical - Analytic Transfer of Training Design

Description	Treatments	Comments
1. Empirical transfer: Control vs. confounded CAI and SIMCAT treatment	Control: Classroom to STTX: N = 12 - 16. To avoid withholding treatments from control and measure order effects, after STTX, half control takes CAI to SIMCAT and half control takes SIMCAT to CAI. Treatment: CAI and SIMCAT to STTX: N = 12 - 16. To test order effects 1/2 takes CAI to SIMCAT to STTX: 1/2 takes SIMCAT to CAI to STTX.	The confounded treatment design does not give transfer for CAI and SIMCAT separately. Cost effectiveness analysis requires separate estimates of effectiveness. If CAI is considered a part-task trainer encompassing enabling objectives, it should transfer to SIMCAT. However, since SIMCAT is a free-play exercise, it is not clear that this transfer can be reliably measured. Direct observation of the treatment order may be appropriate. However, it may not make any difference in what order CAI and SIMCAT are administered
2. Analytic interpolation of separate effects of CAI and SIMCAT	Given the results from number 1 above, the confounded empirical study, SME's extrapolate transfer results for a full year and interpolate the separate effects of CAI and SIMCAT with regard to performance on the STTX. Judgmental variances are method variances (CAI, SIMCAT order effects) student variances and instructor variances on STTX. SME's are as in study 1: 3-5	Extrapolation to full year results projects a long term estimate for the confounded treatments. Interpolation gives estimates of the separate effects of CAI and SIMCAT. Both sets of data will be used as predictors for the empirical results in 3 below.

Table 21 (con't)

Description	Treatments	Comments
	ARI, TTFA and contractor; 3-5 instructors experienced with the methods.	
3. Lagged empirical treatments: CAI only and SIMCAT only.	Treatment 3.1-CAI to STTX; N = 12 - 16. Order effects tested by giving SIMCAT after STTX.	Analytic data is correlated with empirical results and accuracy of estimates determined.
	Treatment 3.2-SIMCAT to STTX: N = 12 - 16. Order effects tested by giving CAI after STTX.	

N for empirical studies - 48 - 64

Transfer within the course may show variance common to a TD/S and WS exercise but would not show the unique variance contributing to battle readiness.

Under certain conditions transfer studies for operator personnel (i.e., tank commanders or gunners) could be conducted after school training in conjunction with a realistic battle exercise, the battle exercise serving as the closest available approximation for a measure of battle readiness. The study would have to compare TD/S transfer and control groups. The conditions that would have to apply to make a transfer study of this type of value would be the following:

1. The battle exercise may be expected to use the skills taught on the TD/S.

2. Training on the TD/S for the transfer group would have to take place shortly before the exercise. Checks would have to be made to assure that the control group had not received training on the TD/S. These are necessary controls for an experiment.

3. TD/S performance can be measured reliably.

4. Performance measures in the battle exercise (i.e., communications flow, hits, kills) are sufficiently reliable to detect a difference between the transfer and control group.

5. There is a sufficient sample size to enable valid statistical tests to be made of the significance of differences and to covary or control for differences among trainees.

6. Adequate attention is given in the experiment to controlling other threats to validity.

Experiments of this type would be useful to relate training and battle readiness, validate judgmental scales of battle readiness, and assess the relative cost effectiveness of a work sample TD/S vs. a battle exercise.

The analytic to empirical study design would use the Job Readiness Scale presented in Chapter 3 to correlate with an empirical transfer study. The empirical study would require a realistic battle exercise that can be scored appropriately to reflect dimensions relevant to SIMCAT and the STTX. Such measures would include command, communication and control variables as well as hits and kills. A variety of covariates would be employed to increase the sensitivity of the experiment. Shortly before the battle exercise treatment groups would participate in SIMCAT and/or STTX and the following experiment carried out:

1. Control group

2. Treatment 1 - SIMCAT only

3. Treatment 2 - STTX only

4. Treatment 3 - SIMCAT and STTX

Comparison of the treatment groups vs. the control group will show the effect of any type of training. Comparison of treatment 3 vs. treatments 1 and 2 will show the common vs. unique contribution of each. The analytic estimates will then be correlated with the empirical data to determine the extent to which they discriminate unique vs. common variance.

The results of the study would also be useful in showing the efficacy of SIMCAT and the STTX for field skill maintenance training for TCs who completed training before SIMCAT and the STTX were available.

CAI lessons could also be added as a treatment group if warranted by earlier transfer study results. However, CAI is considered a part-task trainer, teaching enabling knowledges and skills rather than providing integrated practice. Because of the complex causal relationships, addition of CAI as another treatment group might unnecessarily complicate the study design and interpretation of results.

Study 5. Cost analysis and cost effectiveness analysis.

The Operating Cost/Hour (OCH) will be obtained as the basic cost measure for CAI lessons (categorized as those designed for exportable common military skills and those specific to TC training to take account of differences in scale of use), SIMCAT TC courseware, and the tank in the STTX. The following Operating Cost Ratios (OCRs) will be obtained:

1. CAI - common military skills hourly costs divided by tank hourly costs.

2. CAI - TC specific courseware hourly costs divided by tank hourly costs.

3. SIMCAT hourly costs divided by tank hourly costs.

OCRs of less than 1.0 will indicate that the TD/S is less costly to operate than the tank. Favorable cost ratios are expected since the M1 Abrams tank is expensive to operate while moving. However, the exact value of the OCRs is unknown. Annualized extensions of the hourly costs will

be made. The costing model is detailed in Volume II of this report.

Effectiveness dimensions will be iterated as the analytic and empirical data matures. Recalling the TD/S effectiveness formula in Chapters 1 and 3:

$$\text{TD/S} \quad E \quad (f) = \left\{ \frac{\text{ToT}, \quad U, \quad JR}{\text{Acq}} \right\} \text{UR},$$

the effectiveness of CAI and SIMCAT will depend on their in-course transfer (ToT), contributions to job readiness (JR), the Utilization Ratios (UR) and acquisition (Acq). Safety is not a relevant element for these TD/S. As data become available, the effectiveness dimensions will be updated and analyzed. In-course transfer of training (ToT) will be combined with battle readiness (JR) using a Multi-Altribute Utility Assessment Method described in Chapter 3.

The expectations of the cost-effectiveness relationship are characterized by the decision chart in Figure 5.

		Effectiveness		
		Less	Same	More
	Less	?	+X	+X
Cost	Same	-	?	+
	More	-	-	?

Adopt +
Reject -
Uncertain ?
Expectancy for CAI and SIMCAT X

Figure 5. Decision Diagram for Evaluating Cost Effectiveness of a TD/S

CAI and SIMCAT are hypothesized to be less costly to operate per hour than the STTX on the tank and to provide effectiveness which is at least equal in certain respects to the STTX. However, since the effectiveness function is not monetized, a production function in monetary terms cannot be expressed. The value of the amount of transfer effectiveness (ToT, JR and UR) may be expressed by MAUM methods to combine the various effectiveness elements and weigh them in relation to costs. The nature of the expected

relationship is shown in Figure 6. The lower the OCR and the higher the MAUM effectiveness, the greater is its value. This graphic can be used to display results for methods (CAI lessons for common military skills, CAI lessons specific to TC BNCOC, SIMCAT) and for data elements (ToT, JR, and UR) as well as an overall representation of cost effectiveness.

It should be noted that effectiveness dimensions are multi-variate and each element should be related to costs as data become available. One would hope that the effectiveness results would all point in the same direction, however contrary results need to be evaluated.

SUMMARY

This chapter has presented a number of general research strategies for TECIT and a validation plan for application to the Tank Commander BNCOC at the Ft. Knox Armor School. Research and validation of the model addresses concept and design phase issues as well as those for a fielded TD/S. Cross-sectional as well as longitudinal approaches are outlined.

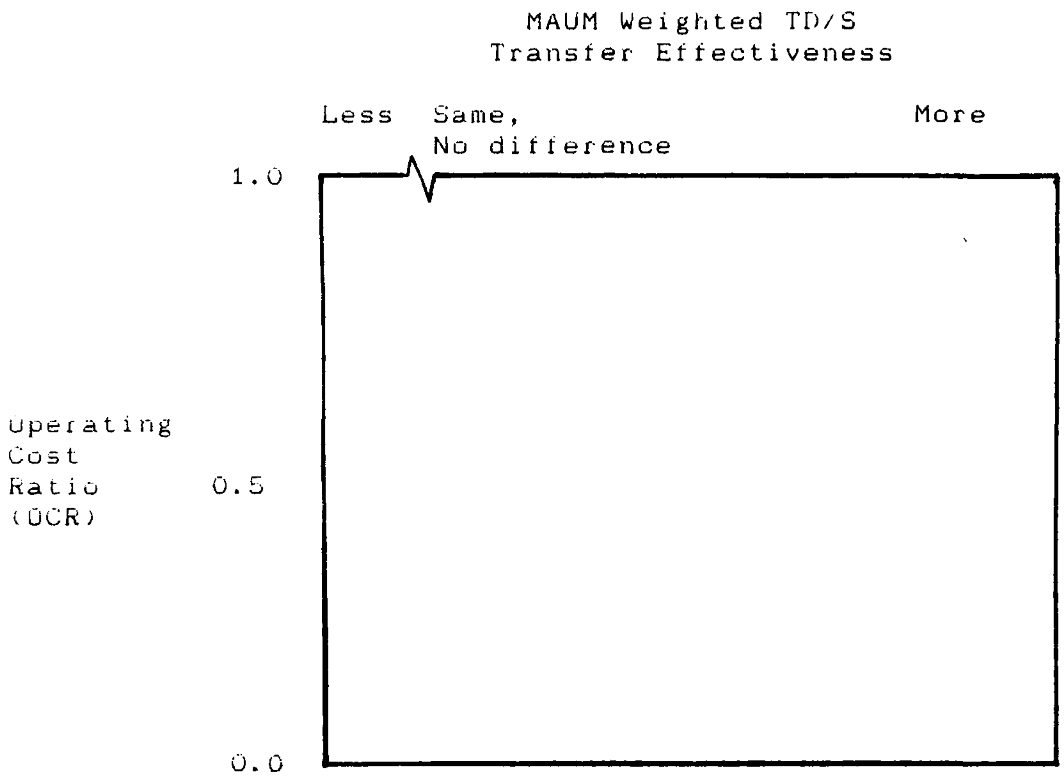

Figure 6: Operating Cost Ratio and Transfer Effectiveness Weighted by the Multi-Attribute Utility Assessment Method (MAUM)

References Cited

Adams, A.V., & Rayhawk, M. (Feb. 1986). A review of models of cost and training effectiveness analysis, Vol. II. cost models. Prepared by the Consortium of Washington Area Universities, the George Washington University, for the Army Research Insititute for the Behavorial and Social Sciences, Alexandria, Va. ARI Research Note 87-59. AD A189645

Blaiwes, A.S., & Regan J.J. (1986). "Training devices: Concepts and progress". Ch. 5 in Ellis, S.A. (ed.), MilitaryContributions to Instructional Technology, Praeger Publishers, N.Y.

Dawdy, E.D. & Hawley, J., A Forecasting Method for Training Effectiveness Analysis. Proceedings of the Human Factors Soiety 26th Annual Meeting, 1982. p. 250-254.

Goldberg, I. & Khattri, N. (Feb. 1986). A review of models of cost and training effectiveness analysis (CTEA)-Volume I: Training effectiveness analysis. Prepared by the Consortium of Washington Area Universities, Univ. of the District of Columbia, for the Army Research Insititute for the Behavioral and Social Sciences, Alexandria, Va. ARI Research Note 87-58. AD A189198

Klein Associates (June, 1985). Comparison-based prediction of cost and effectiveness of training devices: A guidebook. Prepared for the Army Research Institute for Behavioral and Social Sciences, Alexandria, Va. ARI Research Note 85-29. AD A170941

Knerr, C.M.; Nadler, L.; & Dowell, S. (Jan., 1985). Training transfer and effectiveness models. Prepared by the Human Resources Research Organization for the Army Research Institute for the Behavioral and Social Sciences, Alexandria, Va. (In Preparation)

Orlansky, J. & String, J. (April, 1979). Cost effectiveness of computer-based instruction in military training. Final Report prepared by the Institute for Defense Analysis for the Secretary of Defense for Research and Engineering.

Orlansky, J. & String, J. (August, 1981). Cost effectiveness of maintenance simulators for military training. Prepared by the Institute for Defense Analysis for the Office of Under-Secretary of Defense for Research and Engineering.

Orlansky, J. & String, J. (Aug., 1977). Cost effectiveness of flight simulators for military training. Vol. I. Use and effectiveness of flight simulators, prepared by the Institute for Defense Analysis for the Office of Director of Defense Research and Engineering.

Orlansky, J. (January, 1985). The cost effectiveness of military training, presentation to the NATO Symposium on Military Value & Cost Effectiveness of Training. January, 1985.

Pfeiffer, M.G.; Evans, R.M.; & Ford, L.H. (Jan. 1985). Modeling field evaluations of aviation trainers. Naval Training Equipment Center, Orlando, FL.

Pfeiffer, M.A., & Scott, P.A. (Dec, 1985). Experimental and analytic evaluation of the effects of visual and motion simulation in SH-3 helicopter training. Naval Training Systems Center, Orlando, Fl.

Povenmire, H.K. & Roscoe, S.N. "Incremental transfer effectiveness of a ground based general aviation trainer". Human Factors, 534-542, 15 (b), 1973.

Rose, A.M. & Wheaton, G.R. (Dec., 1984). Forecasting device effectiveness: I. Issues, Technical report. Prepared by the American Institutes for Research for the Army Research Institute for the Behavioral and Social Sciences, Alexandria, Va. ARI Technical Report 680. AD A159576.

Rose, A.M. & Wheaton, G.R. (Dec., 1984) Forecasting device effectiveness: II, Procedures. Prepared by the American Institute for Research for the Army Research Institute for the Behavioral and Social Sciences. ARI Research Product 85-25. AD A159955.

Rose, A.M. & Martin, A.M. (March, 1981). Forecasting device effectiveness: III, Analytic assessment of Device effectiveness forecasting technique, Prepared by the American Institutes for Research for the Army Research Institute for the Behavioral and Social Sciences. Alexandria, Va. ARI Technical Report 681. AD A160029.

Tufano, D.R., & Evans, R.A. (April, 1982). The prediction of training device effectiveness: A review of Army models. Technical report of the Army Research Institute for Behavioral and Social Sciences. ARI Technical Report 613. AD A146937.

Wheaton, G.R.; Fingerman, P.; Rose, A.; & Leonard, R. (July, 1976). Evaluation of the effectiveness of training devices: Elaboration and application of the predictive model. Research memorandum 76-16. Prepared by the American Institutes for Research for the Army Research Institute for the Behavioral and Social Sciences, Alexandria, Va. AD A076818.

Appendix A

A Sample Data Base: Orlansky and String's Data on Flight Simulators, Maintenance Simulators and Computer Based Instruction

In a series of reports in 1977, 1979, and 1985, Orlansky and String compiled results from empirical acquisition and transfer of training studies for flight simulators (34 studies), maintenance simulators (13 studies), and Computer-Based Instruction (CBI, 40 studies). These data are presented here to illustrate how a database might prove useful in an analytic model of TD/S. When empirical studies are available comparisons of newly proposed TD/S can be made with similar TD/S in the database. Cost and cost effectiveness data were presented in the original reports, when available, but are not presented here. It should be noted that the types of TD/S are limited to flight simulators and maintenance simulators, limiting the generality of the results to other weapon systems or jobs (i.e., tanks, gunnery). It is also noteworthy that maintenance simulators and CBI used acquisition learning compared with standard classroom instruction as the basis for comparison while transfer of training was used to evaluate flight simulators. Transfer data were generally not available for maintenance simulators or CBI. There were no studies reported that used performance transfer measures as opposed to the time to criterion transfer measures used with flight simulators. The expansion and compilation of empirical studies would be helpful in compiling broader data bases.

Flight Simulators

Figure 1. The frequency distribution is shown of 34 Transfer Effectiveness Ratios (TER) for flight simulators calculated from 22 studies conducted during 1967-1977. The median TER was 0.48. The TERs ranged from -0.4 to 1.9. See Chapter 3 for the TER formula.

Figure 2. The frequency distribution is shown of the Percent Time Saved (PTS) from the simulator to the aircraft in the same studies reported in Figure 1. See Chapter 3 for the PTS formula. The median PTS value was 41% with a range of -9% to +90%.

Figure 3. The relationship is shown of the Transfer Effectiveness Ratio to Percent Time Saved for 31 studies reported in Figures 1 and 2. The two measures are correlated 0.49. The codes for each data point facilitate reference to detailed data in Table 1.

Table 1: This table presents the descriptive and quantitative data available on the various studies done on the flight simulators. Particular simulator characteristics may be useful for comparison.

Table 2: Contains data on the TEA studies done on simulators. Time savings, number of students used and achievement in the various studies are tabulated.

Figure 4. The TER's for 24 maneuvers on which 24 pilots were trained in the Ch-47 helicopter flight simulator is shown. The TERs ranged from 0.00 to 2.8, which suggests that the simulator was effective for those maneuvers with high TERs, i.e., cockpit run up, but not for those with low TERs, i.e., pinnacle approach (Orlansky & String, 1985).

Maintenance Simulators

Figure 5. The results are given of 13 studies conducted during 1967-1980 on the effectiveness of maintenance simulators. Comparisons of end-of-course test scores showed that in 12 cases, students using simulators showed the same or better performance. For 1 case, the scores were lower for the students using the simulator. The differences, though statistically significant, were small. Time saved by students-on-simulators was reported in three studies. These showed that 22%, 50% and 50% of the time needed to complete the course was saved by students-on-simulators as compared to students on the actual equipment. Attitude surveys showed that in nine out of ten cases the students favored the use of simulators, while instructors were equally divided in being favorable, unfavorable and neutral.

Computer-Based Instruction (CBI)

Figure 6. A total of 40 studies are shown here comparing student achievement for CBI and conventional instruction. Of the 40 studies comparing achievement, 1 found CBI to be inferior, 24 to be the same, and 15 to be superior than conventional instruction.

Figure 7. The amount of student time saved by CBI compared with conventional instruction is shown. The results are reported as percent of time saved by CBI.

$$\text{Percent of time saved} = \frac{\text{Conventional - CBI}}{\text{Conventional}} \times 100$$

The median time saved was 30% ranging from -31% to 80%.

Figure 8. In 12 courses, Individualized Instruction (programmed instruction) and CAI or CMI were compared with conventional instruction for student time savings. In five courses, Individualized Instruction saved an average of 64% of student time and CAI saved an average of 69% of student time. In seven courses both individualized instruction and CMI saved an average of 51% of student time.

Figure 9. The actual student time saving with individualized, CAI and CMI instruction as compared with conventional instruction in the same courses are shown. The range is from 30% to 90% savings with no statistically significant differences between Individualized Instruction and CAI or CMI.

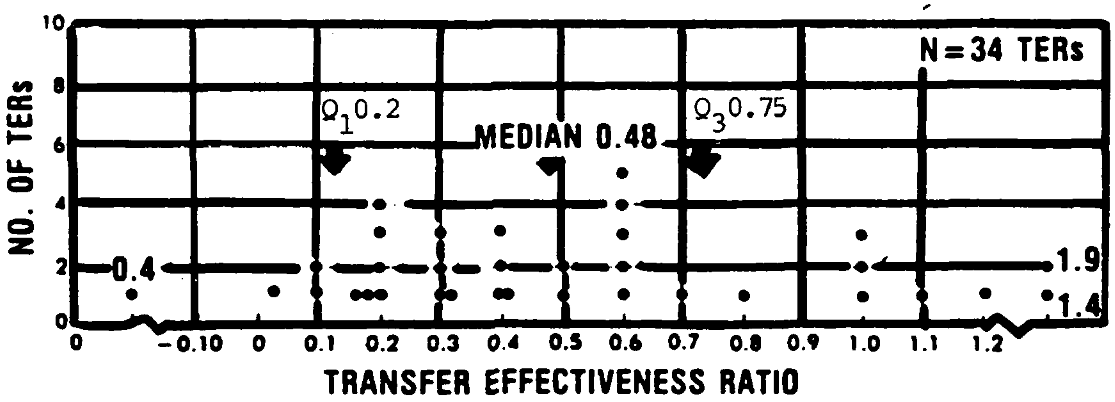

FIGURE 1: Transfer Effectiveness Ratios of Flight Simulators 22 Studies (1967-1977)

SOURCE: Orlansky & String, 1985.

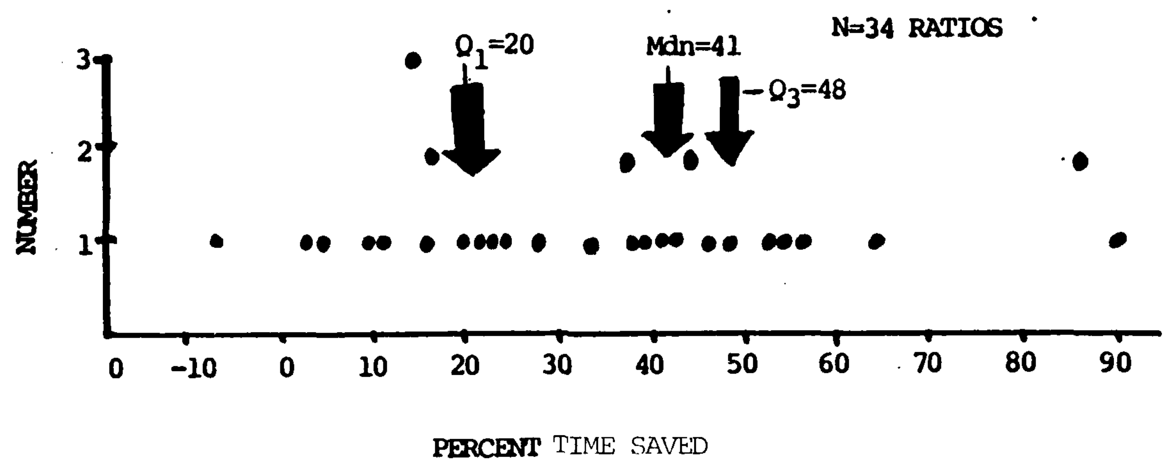

FIGURE 2: **Percent** Time Saved of Flight Simulators, 22 studies (1967-1977)

SOURCE: Tabulated from Orlansky & String (1977), Table 5.

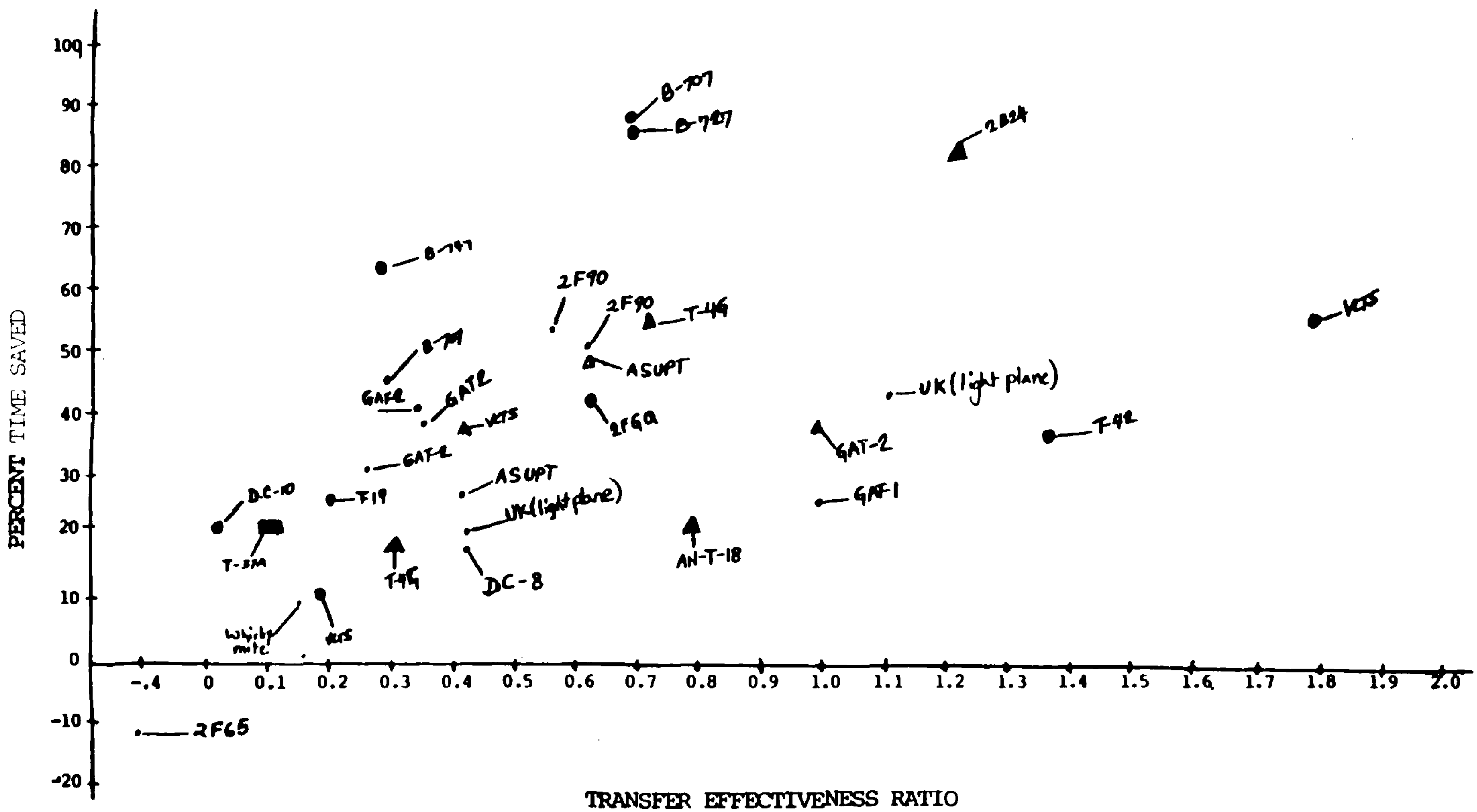

FIGURE 3: Relationship of Transfer of Training Measures—Transfer Effectiveness Ratio and Percent Time Saved

SOURCE: Tabulated from Orlansky and String, (1977), Table 5

CURRICULUM FEATURES

- • = Not known
- ▲ = Special Syllabus
- ● = Special Syllabus and part task trainer
- ■ = Part task trainer
- Δ = Basic and Presolo

TABLE 1 TRANSFER OF TRAINING MEASURES, CALCULATED FROM REPORTED INFORMATION

Tasks	Aircraft	Simulator	Student* Experience	Simulator Capabilities	Curriculum Features	Percent Transfer	Transfer Effective-ness Ratio	Flight Substitu-tion Ratio	References
Contact/ Familiar-ization	Piper Cherokee	AN-T-18	Undergraduate	-	Special Syllabus	20	0.8	1.2	Povenmire and Roscoe (1971)
Contact/ Familiar-ization	Piper Cherokee	GAT-1	Undergraduate	-	Special Syllabus	24	1.0	1.0	Povenmire and Roscoe ([illegible])
Contact/ Familiar-ization	Light Plane	?	Undergraduate	-	-	16	0.4	2.3	Crook (1967)
Instru-ments	Light Plane	?	Graduate	-	-	48	1.1	0.9	Crook (1967)
Transi-tion	B-747	B-747	Highly Experienced	Visual/ Motion	Special Syllabus Part-Task Trainer	64	0.2	-2.6	Melden and Houston (1975)
Transi-tion	B-707	B-707	Highly Experienced	Visual/ Motion	Special Syllabus Part-Task Trainer	90	0.6	-0.7	" " "
Transi-tion	B-727	B-727	Highly Experienced	Visual/ Motion	Special Syllabus Part-Task Trainer	89	0.6	-0.9	" " "
Transi-tion	DC-10	DC-10	Highly Experienced	Visual/ Motion	Special Syllabus Part-Task Trainer	23	0.03	-8.0	" " "
Instru-ments	UH-1	2B24	Undergraduate	Motion	Special Syllabus	89	1.2	0.8	Caro (1973)

*Undergraduate refers to military UPT programs and general aviation student pilot training programs
Graduate refers to designated military pilots and licensed general aviation pilots
Highly Experienced refers to airline pilots
Naive refers to no previous flight experience
Sources: From Orlansky and String (1977), Diehl and Ryan (1977), Micheli (1972) and original sources.

TABLE 1 (Continued)

Tasks	Aircraft	Simulator	Student* Experience	Simulator Capabilities	Curriculum Features	Percent Transfer	Transfer Effectiveness Ratio	Flight Substitution Ratio	References
Qualifications	H-52	VCTS	Graduate	Motion	Special Syllabus Part-Task Trainer	54	1.9	0.5	Isley, Corley and Caro (1974)
Transition	H-52	VCTS	Graduate	Motion	Special Syllabus Part-Task Trainer	10	0.2	5.8	Isley, Corley and Caro (1974)
Transition	H-3	VCTS	Graduate	Motion	Special Syllabus	36	0.4	2.3	Isley, Corley and Caro (1974)
Transition	H-3	T-42	Graduate	Motion	Special Syllabus Part-Task Trainer	41	1.4	0.7	USAF (1974)
Contact/Familiarization	T-37	T-4G	Undergraduate	Visual/Motion	Special Syllabus	15	0.3	3.7	Woodruff, Smith and Morris (1974)
Instruments	T-37	T-4G	Undergraduate	Visual/Motion	Special Syllabus	53	0.7	-0.8	Woodruff, Smith and Morris (1974)
Familiarization, Instrument Flight	T-37	ASUPT	Undergraduate	Visual/Motion	Basic and Presolo	45	0.6	-	Woodruff, Smith, Fuller and Weyer (1976)
					Advanced Contact	4	0.1	-	
					Instruments	38	0.5	-	
					Formation	13	1.0	-	
					Navigation	13	0.2	-	
					Total	23	0.4	2.1	
Instruments	TA-4	2F90	Undergraduate	Motion	-	52	0.5	0.4	Ryan, Puig, Micheli and Clark (1972)
Instruments	TA-4	2F90	Undergraduate	Motion	-	46	0.6	0.4	O'Connor and Glennon (1973)
Instruments	TA-4	2F90	Undergraduate	Motion	-	01	-	42.0	O'Connor and Glennon (1973)
Familiarization, Instrument Flight	P-3	2F69	Graduate	Motion	Special Syllabus Part-Task Trainer	39	0.6	0.3	Browning, Ryan and Scott (1973)

TABLE 1 (Continued)

Tasks	Aircraft	Simulator	Student* Experience	Simulator Capabilities	Curriculum Features	Percent Transfer	Transfer Effectiveness Ratio	Flight Substitution Ratio	References
Familiarization, Instrument Flight	P-3	2F87F	Graduate	Visual/ Motion	-	43	0.3	2.3	Browning, Ryan, Scott and Smode (1977)
Familiarization, Instrument Flight	E-2	2F65	Graduate	-	-	-11	-0.4	-	Diehl and Ryan, Ref. 22 (1977)
Familiarization, Instrument Flight	C-130	T-19	Graduate	Motion	Special Syllabus Part-Task Trainer	22	0.2	0.4	Diehl and Ryan, Ref. 23 (1977)
Familiarization, Instrument Flight	C-141	T-37A	Graduate	Visual/ Motion	Part-Task Trainer	15	0.1	15.4	Diehl and Ryan, Ref. 24 (1977)
Familiarization, Instrument Flight	T-42	GAT-2	Undergraduate	Motion	Special Syllabus	42	1.0	0.2	Caro, Isley and Jolley (1973)
Procedures; Takeoff; Hover, Landing	Helicopter	Whirlymite	-	-	-	9	0.17	-	Caro (1968)
Flight Procedures Maneuvers	B-707	B-707	Highly Experienced	-	-	49	0.19	-	TWA Training Dept. (1969)
Flight Procedures and Maneuvers	DC-8	DC-8	Highly Experienced	-	-	13	0.41	-	Meyer, Flexman (1967) Van Gundy, Killiam and Lanahan
Private Pilot Certification	Piper Cherokee Arrow	GAT-2	Naive	No motion Motion Random Motion	Highly Standardized	36 34 28	0.30 0.31 0.25	3.2 3.3 4.0	Jacobs and Roscoe (1975)

TABLE 1 (Continued)

Tasks	Aircraft	Simulator	Student* Experience	Simulator Capabilities	Curriculum Features	Percent Transfer	Transfer Effectiveness Ratio	Flight Substitution Ratio	References
Night Carrier Landing	A7E	Night Carrier Landing Device 2F103	Pilots with no previous A7E experience (N=53)	Visual $40^{\circ} \times 30^{\circ}$ V Motion 3 DOF	80 final approach control trials in simulator vs none for control group, latter received familiarity training in simulator, objective performance measures by radar on carrier landing.	(1)	-	-	Brictson and Burger (1976)

(1) Radar measures show that simulator *trained pilots* show more precision in vertical flight control than those who did not; attrition rate lower, for newly designated pilots with simulator training (8 percent) than for those who did not get such training (44 percent). Flight time savings not measured.

TABLE 2 SUMMARY OF STUDIES ON THE EFFECTIVENESS OF FLIGHT SIMULATORS

Study	Vehicle	Simulator	Skills Taught	Experiment	Results
Mahler and Bennett (1949) review					
NRC, 1939	Civilian light aircraft	Link AN-T-18	Basic contact flight for civilian pilot training program	Analysis of performance records (N=10). No control group.	Estimate that 5 to 7 hrs. in trainer was equivalent to 3 hrs in A/C. Savings of 2 to 4 hrs air time (Inconclusive)
NRC, 1940	Civilian light aircraft	Link AN-T-18	Basic contact flight for civilian pilot training program.	Analysis of performance records (N=10). No control group.	Estimated 24 hrs saving in air time with 6 hrs of trainer time. (Inconclusive)
NRC, 1941	Civilian light aircraft	Link AN-T-18	Basic contact flight for civilian pilot training program.	Instructor performance ratings. Three groups of 11, 8, and 11 civilian pilot training students	Groups with more Link trainer time were rated higher than group with only one hour of training. (Inconclusive)
Naval Reserve Aviation Base, Long Beach, CA 1942	Military	Link	Basic flight training.	(N=146) No control group	(1) Reduced number of dual instruction hrs for solo. (2) Reduced number of students receiving downs on their check flights. (Inconclusive)
Naval Flight Preparatory School, William Jewell College, Liberty, Missouri 1943	Military	Contact Link	Basic flight training.	(N=1400) 1/2 received 10 one hr sessions on Contact Link Trainer. Other ½ no synthetic training.	Experimental students tended to slight advantage over control students in capability for solo time, actual solo time, and instructor's grades. (Differences were not statistically significant.)

TABLE 2 (Continued)

Study	Vehicle	Simulator	Skills Taught	Experiment	Results
Naval Air Station, Memphis 1945	Military	12BK-1 Primary Landing Trainer	Primary training	(N=166) 1/2 experimental and 1/2 control	(1) Experimental students completed syllabus faster than controls by 10%. (2) Control group had 10% more flight failures in A stage, 5% more in B stage. (3) Differences disappear by end of C stage.
Univ. of Illinois, 1949	SNJ	(1) 12-BK-1 Landing Trainer (2) C-3 Cycloramic Link Trainer (3) SNJ Cycloramic (General) Link (1-CA-2)	Instrument training and control skills	(1) experienced 5 s. (solo flight time) (N=234) (2) 10 hrs syn. trainer (N=405) (3) Control group (N=127)	Three trainers equivalent accidents reduced 40%, failure rate down 33%
Williams and Flexman (1949)	SNJ-5 Modified for civilian use	SNJ Cycloramic Link (1-CA-2)	Basic contact flight training	(2 groups of 6 college students ea. 12 hr [illegible] syllabus.) Trainer group 8 hrs on trainer. Control group 11 hrs. A/C.	12 hrs to learn in air, 5 hrs air time for trainer group. Fewer errors for trainer group.
Mahler and Bennett (1950)	PBM (2-engine seaplane) PB4Y (4-engine land plane)	PBM-OFT PB4Y-OFT	Familiarization and instrument training.	Series of controlled experiments using 23 matched prs. of students for each trainer	Flight time reduced 1½ hrs. out of 12 hr syllabus for INST, no saving for FAM stage. Fewer errors in both stages.

TABLE 2 (Continued)

Study	Vehicle	Simulator	Skills Taught	Experiment	Results
Brown, Matheny and Flexman (1950)	SNJ	School Link with "blackboard" runway.	Ground reference maneuvers (landings, forced landings, pylon 8's)	N=20 college students 10 on trainer 10 principles training	Trainer group = 2.59 errors Control group = 4.29 errors
Wilcoxon, Davy and Webster (1954)	SNJ	SNJ OFT (Specialized electronic high fidelity trainer) and NavBIT (General low fidelity basic instrument and radio range trainer)	Instr. training including radio range.	Progress-at-own-rate syllabus and ground training under a blocked sequence Std. Blk Syl NavBit N=96 OFT N=33 Emp. Black Syl N=168 N=52	(1) saved an av. of 1.3 hrs in flight or ~ 3000 hrs/hr or, 1 flight out of 11 bas. inst. flts. (2) NavBit equal in effectiveness to SNJ OFT for basic instrument training and slightly superior for radio range work.

TABLE 2 (Continued)

Study	Vehicle	Simulator	Skills Taught	Experiment	Results
Flexman, Towsend and Ornstein (1954)	T-6 (Navy SNJ)	P-1 (1-CA-2 SNJ (Cycloramic Link)	Procedures, manuevering	95 aviation cadets, 47 in trainer. Substituted 40 simul. hrs for 30 flt hrs in a 130 hr syllabus	40 sim. hrs = 30 flt hrs ratio = 0.75
Payne, Doherty, Hasler et al (1954)	SNJ	Cycloramic Link	Approach and landing	Experimental group (N=6) vs control group (N=6).	61% fewer trials & 74% fewer errors for simulator group.
Dougherty, Houston and Nicklas (1957)	SNJ	Cycloramic Link, photo-mockup, procedures trainer	Procedures	3 trainer groups compared to each other and flight group.	All groups equal after three air trials.
Isley, Caro and Jolley (1968)	Army Helicopter	1-CA-1 modified to rotary-wing configuration.	Instrument flight in rotary wing training. (U.S. Army Aviation School)	Total N=145, 3 groups: 0 hrs, 10 hr, and 20 hr synthetic training. All groups received 25 hrs flight training.	No significant difference between groups.
Caro, Isley and Jolley (1968)	Army Helicopter	"Whirlymite" captive helicopter	Rotary wing contact flight.	Total N=132. Divided into 2 experimental groups and 2 control groups with no training on device. 0, 3½, 7½ hrs of practice.	(1) 10% attrition for flying deficiencies in exper. groups. 30% attrition for control groups. (2) Two hrs less flight training needed to solo for exper. groups. (3) Flight grades higher early in training.

TABLE 2 (Continued)

Study	Vehicle	Simulator	Skills Taught	Experiment	Results
Crook (1967)	light aircraft	ground trainer	private pilot course, instrument flight		16 sim. hrs reduce flight from 42 to 35 hrs. 17 sim. hrs reduce flight from 40 to 21 hrs.
TWA (1969)	B707	B707	Flight procedures		
Houston (1970)	B727 BAC 400	B727 BAC 400	Flight procedures Flight procedures		
Caro, Isley and Jolley (1973)	T-42	GAT-2	transition and instrument qual.		Increase from 21 hrs in old sim. to 25 hrs in new sim reduced flight from 60 to 35 hrs.
Caro (1972)	UH-1 helicopter	2B24	Instrument flight		43 sim hrs reduce flight from 60 to 7 hrs.
Browning, Ryan and Scott (1973)	P-3	2F69	four engine turboprop transition course		increase from 11 to 14 sim hrs reduced flight from 19 to 12 hrs.
O'Connor and Glennon (1973)	TA-4	2F90	Basic instrument and navigation		revised course increases sim from 21 to 27 hrs and reduces flight from 35 to 19 hrs

TABLE 2. (Continued)

Study	Vehicle	Simulator	Skills Taught	Experiment	Results
Robins and Finley (1972)	P-3A	2F69B (P-3A Weapon System Trainer)	Air ASW Tactics	Practice in trainer, then transfer to flight	Improved on the trainer, in-flight transfer data analysis not yet completed.
Ryan, Puig, Micheli and Clarke (1972)	TA-4J	2F90 (TA-4J Operational Flight Trainer)	Basic Instrument Flight	Groups: (1) Control (standard training) (2) Flight only (3) Simulator only (4) Academic only 30-33 subjects per group	No difference on flight check between flight and trainer groups. Simulator group saved 4 flight hours (55%) (1)
Woodruff, Smith and Morris (1974)	T-37	T-4G	Contact flight		15 sim hrs reduce flight from 27 to 23 hrs
	"	"	Undergraduate instrument flight		Revised course reduces sim hrs from 23 to 15 and flight from 21 to 10 hrs.

(1) Report says "The substitution of trainer time for time on the operational equipment is an excellent way of increasing cost-effectiveness" (p.1) but no cost data are provided.

TABLE 2 (Continued)

Study	Vehicle	Simulator	Skills Taught	Experiment	Results
Isley, Corley and Caro (1974)	H-52 helicopter	VCTS	Search and rescue qualifications		22 sim hrs reduce flight from 78 to 36 hrs
	"	"	Search and rescue transition		18 sim hrs reduce flight from 31 to 28 hrs
	H-3 helicopter	"	" " "		30 sim hrs reduce flight from 36 to 23 hrs
A/S Rescue and Recovery Service (1974)	H-3 helicopter	T-42	Transition to search and rescue		19 sim hrs reduce flight from 63 to 37 hrs
Woodruff, Smith, Fuller and Weyer (1976)	T-37	ASUPT	Basic jet course		Revised course increases sim from 17 to 60 hrs and reduces flight from 91 to 71 hrs
Brown (1976) Brown (1975)	B-747	B-747	Transition training		sim. 28-19, a/c 6-2
	B-707	B-707	" "		27-19 13-1+
	B-727	B-727	" "		28-19 12-1+
	DC-10	DC-10	" "		23-19 2-2
Diehl and Ryan, (1977) p.21	E-2	2F65	2 engine turboprop transition course		15 sim hrs (compared to none) increased flight from 61 to 68 hrs

TABLE 2 (Continued)

Study	Vehicle	Simulator	Skills Taught	Experiment	Results
Browning, Ryan, Scott and Smode (1977)	P-3	2F87F 2F69D	four engine turboprop transition course for ASW	27 pilots trained on 2F87F and P-3C and new curriculum compared to 16 trained on 2F69D and P-3C on old curriculum	increase from 9 hrs in old sim to 24 hrs in new sim reduces flight from 15 to 9 hrs.
Eddowes (1977)	T-37	ASPT	Undergraduate maneuvers		
	T-37	T-4G	contact flight		
	T-37	T-4G	instrument flight		
	T-37	T-4	" "		
Diehl and Ryan, p.22 (1977)	C-130	T-19	transition to 4 eng. turboprop		increase in sim from 30 to 32 hrs reduced flight from 23 to 18 hrs
" "	C-141	T-37A	" " "		40 sim hrs (from none) reduced flight from 17 to 15 hrs

MANEUVER	TER
FOUR WHEEL TAXI	2.80
COCKPIT RUN UP	1.50
SAS OFF FLIGHT	1.33
DECELERATION	1.25
MAXIMUM TAKE OFF	1.25
GENERAL AIR WORK	1.00
STEEP APPROACH	1.00
TWO WHEEL TAXI	1.00
CONFINED AREA RECON	1.00
HOVERING FLIGHT	0.79
NORMAL TAKE OFF	0.75
CONFINED AREA APPROACH	0.75
LANDING FROM HOVER	0.69
EXTERNAL LOAD BRIEFING	0.67
TAKE OFF TO HOVER	0.63
TRAFFIC PATTERN	0.61
SHALLOW APPROACH	0.58
NORMAL APPROACH	0.53
CONFINED AREA TAKE OFF	0.50
EXTERNAL LOAD TAKE OFF	0.50
EXTERNAL LOAD APPROACH	0.50
PINNACLE RECON	0.50
PINNACLE TAKE OFF	0.33
PINNACLE APPROACH	0.00

Source: Holman, G.L., 1979.

FIGURE 4 : Transfer Effectiveness Ratios, 24 maneuvers, CH-47 Flight Simulator (Trials to Criterion)

SOURCE: Orlansky & String, (1985)

SIMULATOR	COURSE	COURSE LENGTH (STANDARD)	NO. OF SUBJECTS	COMPARISONS: SIMULATOR TO ACTUAL EQUIPMENT					
				EFFECTIVENESS				ATTITUDE TO SIMULATORS	
				POORER	SAME	BETTER	TIME SAVINGS	STUDENTS	INSTR.
Generalized Sonar Maintenance Trainer	Sonar maintenance (special course)	4 days	9		•		22%	+	
	Intermediate General Electronics	4 weeks	20		•				
EC II	APQ-126 Radar		17					+	
	Mohawk Propeller System	3 hrs	33			•			
	Hydraulic and Flight Control	32 hrs	13		•	•		+	–
	Engine, Power Plants and Fuel	24 hrs	13	•	•			+	–
	Environmental/Utility System	32 hrs	9		••			–	
	APQ-126 Radar	60 hrs	15		••			0/+	0/–
								+	–
	Pilot Familiarization, T-2C	10 hrs	6						–
	Flight Officer Familiarization, TA-4C	11 hrs	30						–
Generalized Maintenance Training System	SRC-20 UHF Voice Command System		20				ABOUT 50%	+	
	SPA-61 Radar Repeater	16 hrs	10					+	
Fault Identification Simulator	Nagan Automatic Boiler	5 wks	16		•		ABOUT 50%		
6883 Converter/ Flight Control Systems Test Station	F-111 Avionics Maintenance	6 days	56		•			–	0/–

FIGURE 5: Studies on the Effectiveness of Maintenance Simulators, 1967–1980

SOURCE: Orlansky & String, 1985.

METHOD OF INSTRUCTION	SYSTEM	SERVICE	LOCATION	STUDENT ACHIEVEMENT AT SCHOOL (compared to conventional instruction)			TYPE OF TRAINING
				INFERIOR	SAME	SUPERIOR	
CAI	IBM 1500	A	SIGNAL C&S		• • • •	• •	ELECTRONICS
		N	SAN DIEGO			• • • • •	ELECTRICITY
	PLATO IV	A	ABERDEEN		• • •		MACHINIST
		N	SAN DIEGO		• •	• • • •	ELECTRONICS
		N	SAN DIEGO		• •		RECIPE CONVERSION
		N	NORTH ISLAND			•	A/C PANEL OPERATOR
		AF	SHEPPARD		• •	• •	MEDICAL ASSISTANT
		AF	CHANUTE		• • • •		VEHICLE REPAIR
	LTS-3	AF	KEESLER		• •	•	ELECTRONICS
		AF	KEESLER		•		WEATHER
	TICCIT	N	NORTH ISLAND		•		TACTICAL CO ORD. (S 3A)
	IDIOM	N	DAM NECK	•	•		FIRE CONTROL TECHNICIAN
	PLATO IV	N	DAM NECK		• •		FIRE CONTROL TECHNICIAN
				TOTAL 1	24	15	
CMI	NAVY CMI	N	MEMPHIS		• •		AVIATION FAMILIARIZATION
		N	MEMPHIS		• •		AV. MECH. FUNDAMENTALS
	AIS	AF	LOWRY		•		INVENTORY MGMT.
		AF	LOWRY		•		MATERIEL FACILITIES
		AF	LOWRY		•		PREC. MEASURING EQPT.
		AF	LOWRY		•		WEAPONS MECHANIC
				TOTAL 0	8	0	

FIGURE 6: Student Achievement at School, CAI and CMI Compared to Conventional Instruction

SOURCE: Orlansky & String, 1985

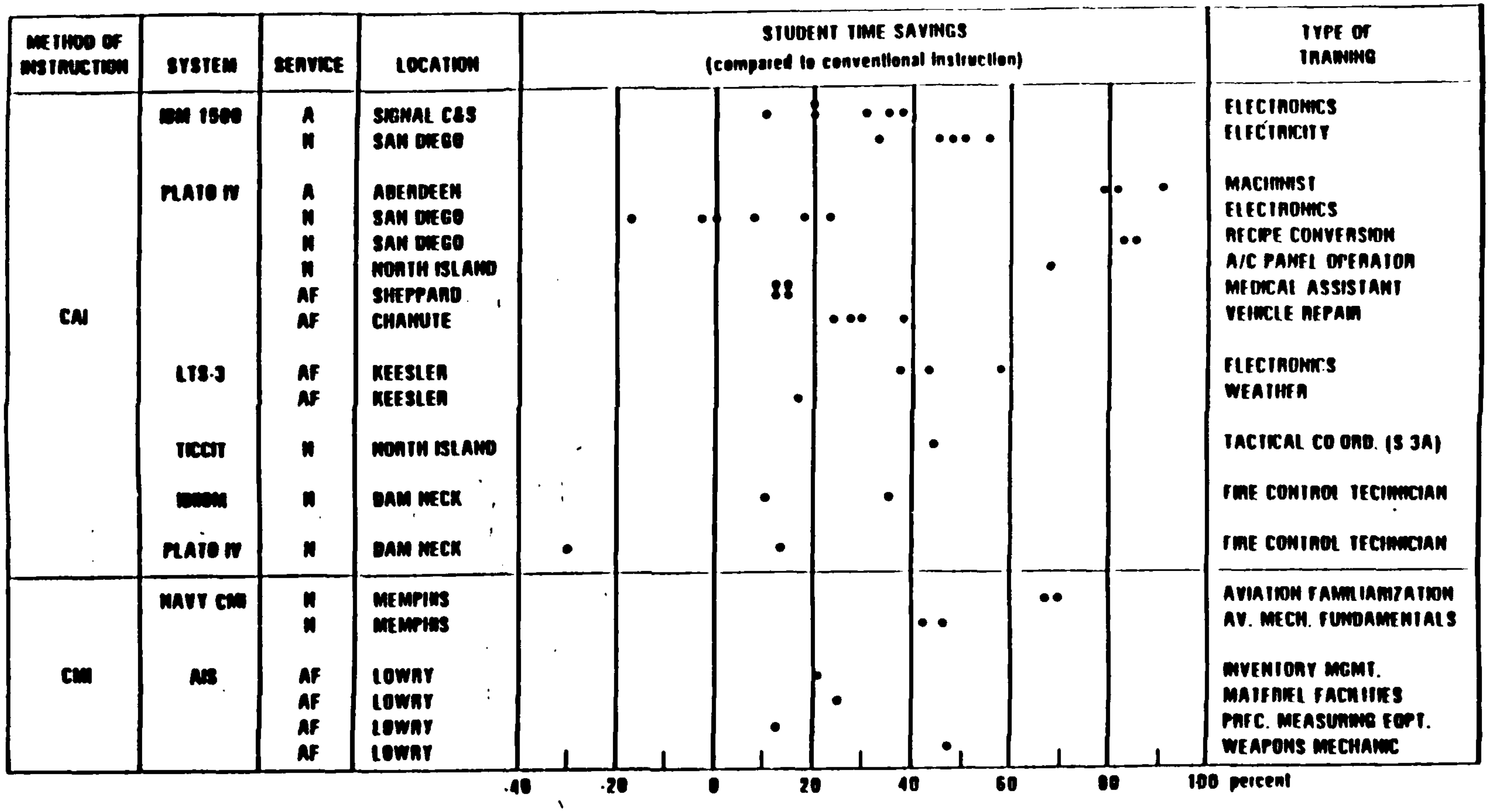

FIGURE 7: Amount of Student Time Saved by CAI and CMI, Compared to Conventional Instruction

SOURCE: Orlansky & String, 1985

NO. OF COURSES	AVERAGE AMOUNT OF STUDENT TIME SAVED		
	INDIVIDUALIZED INSTRUCTION	CAI	CMI
5	64%	69%	—
7	51%	—	51%

FIGURE 8: Average Amount of Student Time Saved by Individualized Instruction and CAI or CMI, Compared to Conventional Instruction

SOURCE: Orlansky & String, 1985

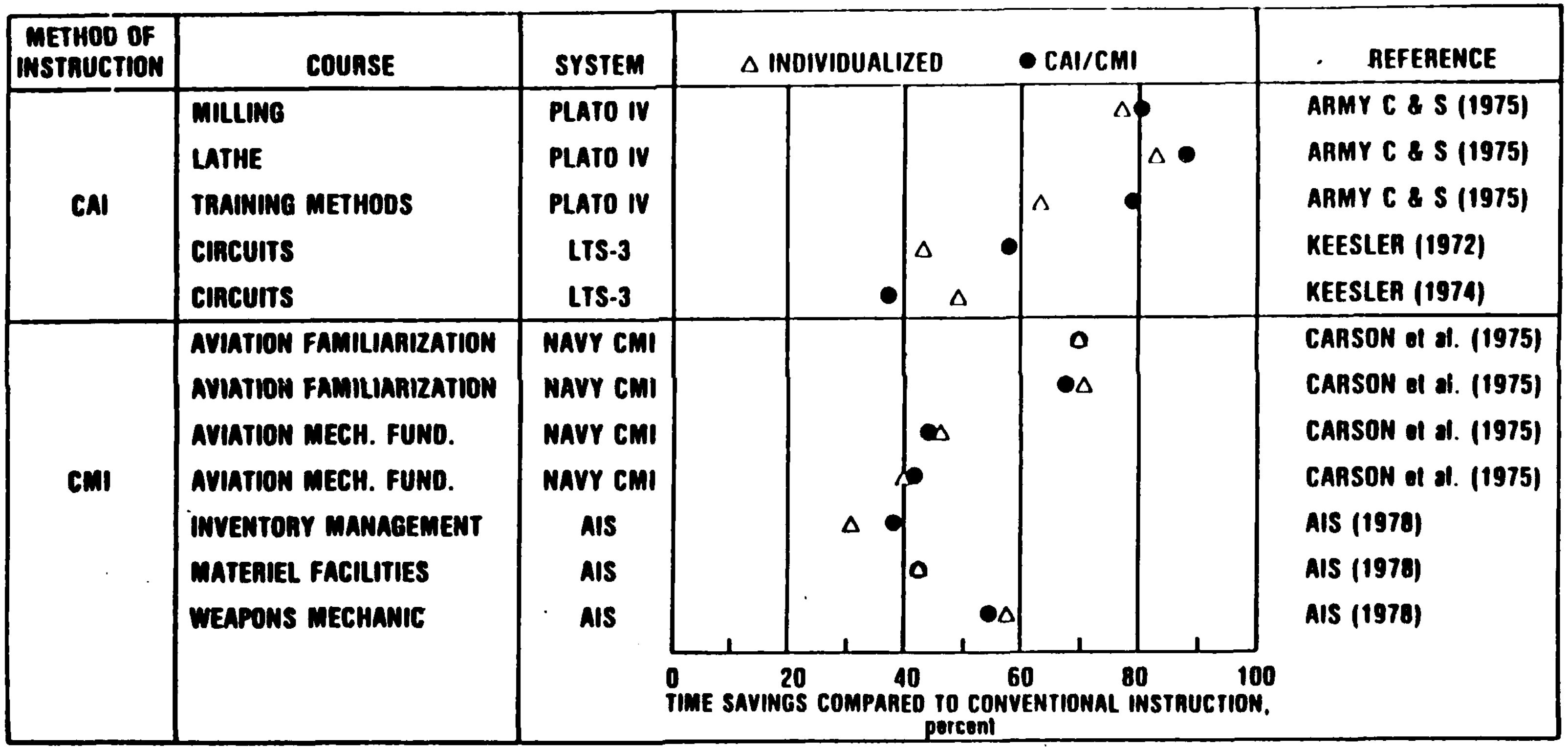

FIGURE 9: Amount of Student Time Saved, Compared to Conventional Instruction, by Individualized Instruction and by CAI or CMI on the Same Courses

SOURCE: Orlansky & String, 1985

Appendix B

Sample Analytic Questionnaires for Transfer of Training Within the Course

1. DEFT I, modified for use in tank training for SIMCAT and a tank exercise (STTX).

2. TECIT I, II, and III. An adaptation of FORTE for use in tank training to measure performance transfer from SIMCAT and Computer Assisted Instruction (CAI) lessons to a tank exercise (STTX).

3. FORTE I & II. Original scales for flight training.

DEVICE EFFECTIVE FORECASTING TECHNIQUE (DEFT)

Training Problem Analysis: DEFT I

PERFORMANCE DEFICIT

I. Examine the statement of the training objective(s). Considering what you know about the typical trainee's background, work experience, and prior training, what proportion of the skills and knowledges required in order to meet the training objective(s) will the trainee still have to learn in order to reach criterion proficiency in SIMCAT?

0 = None; the trainee can already meet the training objective(s).

0

100

100 = All; the trainee has to learn all of the skills and knowledges needed to meet the training objective(s).

LEARNING DIFFICULTY

II. Consider the enabling skills and knowledges required to meet the training objective(s) that the typical trainee does not currently possess. Rate the difficulty of acquiring the remaining skills and knowledges in SIMCAT.

0 = Very easy to learn the skills and knowledges needed to meet the training objective(s) on SIMCAT.

0

100

100 = Very difficult to learn the skills and knowledges needed to meet the training objective(s) on SIMCAT.

Acquisition Efficiency Analysis: DEFT I

QUALITY OF TRAINING ACQUISITION

I. Examine information about the instructional features of SIMCAT, the training principles it incorporates, the program for its implementation, and the larger training context in which it is embedded. Consider the performance deficits you have identified and how utilization of SIMCAT will overcome these deficits.

To provide "excellent" training, the training system should:

- o make the performance requirements of the training objective(s) explicit to the trainees:
- o provide meaningful and understandable feedback to the trainee regarding the results of his performance as soon as possible following his performance:
- o provide sufficient practice where specific and hard-to-learn physical skills are involved; and

Rate the quality of the training provided by this training system, considering only the training problems you have identified.

0 = Poor training; the system embodies few if any sound training principles and instructional features.

0

100

100 = Excellent training; the system makes maximum use of sound training principles and instructional features.

Transfer Problem Analysis DEFT I

RESIDUAL DEFICIT

I. Assume that the trainee has achieved the training objective(s) (i.e., has reached criterion proficiency on SIMCAT. What proportion of the skills and knowledges required in order to reach criterion proficiency on the operational equipment will the trainee still have to learn?

0 = None; the trainee can already meet the operational performance objectives.

0

100

100 = All; the trainee has to learn all of the skills and knowledges needed to meet the operational performance objective(s).

RESIDUAL LEARNING DIFFICULTY

II. Consider the skills and knowledges that a graduate of SIMCAT must still acquire in order to perform at criterion level(s) on the operational equipment. Rate the difficulty of acquiring the remaining skills and knowledges.

0 = Very easy to learn; it will take practically no training on the Tank to learn the skills and knowledges needed to meet the operational performance objectives(s).

0

100

100 = Very difficult to learn; it will take a lot of training on the tank to learn the skills and knowledges needed to meet the operational performance objective(s).

PHYSICAL SIMILARITY

Physical similarity is based on the similarity between physical characteristics of SIMCAT and those of the operational situation. The assessment is based on the physical similarity (e.g., location, appearance, and feel) of displays, controls, and ambient conditions in the training and operational setting. Determine the physical similarity between SIMCAT and the Tank FTX.

0 = Totally dissimilar; there would be a large noticeable difference, quite apparent to the trainee at transfer and a large performance decrement, given that the trainee could perform at all; specific instruction and practice would be required on the operational equipment after transfer to overcome the deficit.

0

100

100 = Identical; the trainee would not notice a difference between the training device and the operational equipment at the time of transfer.

FUNCTIONAL SIMILARITY

Functional similarity is based on the operator's behavior in terms of the information flow from each display to the operator, and from the operator to each control. The assessment is made in terms of the amount of information transmitted from each display to each control and the type of information-processing activity performed by the operator. Determine how functionally similar SIMCAT and the Tank are.

0 = Totally dissimilar; the trainee acts on completely different types and amount of information in SIMCAT and the Tank FTX: the trainee carries out different information-processing activities.

0

100

100 = Identical; the trainee acts on the same types and amounts of information in SIMCAT and the Tank equipment; the trainee carries out the same information-processing activities.

Transfer Efficiency Analysis DEFT I

QUALITY OF TRAINING TRANSFER

I. Consider the statement of the operational performance objective(s), as given in the Training Device Requirement Document , the statement of the training objective(s), performance measure(s) and descriptions of the tank and the SIMCAT excercise.

Consider the instructional features and training principles that are included in SIMCAT to increase the probablility that the skills and knowledges acquired on the device will be used effectively in the operational situation. Rate how well the training device will promote transfer to the operational situation.

0 = Poor transfer; the device embodies few if any sound training principles and instructional features to promote transfer to the operational equipment.

0

100

100 = Excellent transfer; the device makes maximum use of sound training principles and instructional features to promote transfer to the operational equipment.

TRAINING EFFECTIVENESS AND COSTS ITERATIVE TECHNIQUE (TECIT)

OVERVIEW: This questionnaire is designed for tank officers, instructors, and experienced developers of training devices and simulators.

It elicits information that will enable evaluators to forecast and guide the design and execution of transfer of training studies involving tank simulators. We are particularly interested in your estimates of the performance of a student tank commander on a variety of training tasks taught by a variety of instructors both with and without the aid of Computer Assisted Instruction and SIMCAT.

Before proceeding, familiarize yourself with the STTX, SIMCAT excersize, and Computer Assisted Instruction lessons developed for the student tank Commander.

I. First, think of a group of student tank commanders who <u>have completed the SIMCAT exercises prior to the STTX in the tank</u>. Please make estimates of performance (percent of "Go's")on the Tank Commanders STTX, under each of the following eight sets of conditions.

INSTRUCTOR	STUDENT	TASK	PERCENT OF "GO's" ON STTX (TECIT I)
1. Easy	Fast	Easy	
2. Easy	Fast	Tough	
3. Easy	Slow	Easy	
4. Tougn	Fast	Easy	
5. Easy	Slow	Tough	
6. Tougn	Fast	Tough	
7. Tougn	Slow	Easy	
8. Tougn	Slow	Tougn	

9. Now, please rank the following variables for their importance to the estimations you just made:

Rank	Variable
	Instructors
	Students
	Tasks

Administrator: Sum the eight sets of trials recorded above and divide by 8. Insert this mean value (rounded to a whole number) following the symbol "*N*" in questions 10-12. (TECIT II)

10. If an average student achieves *N* "Go's", how many "Go's" will

... a fast learner receive?
... a slow learner receive?

11. If an average instructor gives *N* "Go's" in training students, how many "Go's" will

... an easy instructor give?
... a tough instructor give?

12. If an average task receives *N* "Go's", how many "Go's" would

... an easy task give?
... a tough task give?

II. Second, think of a group of student tank commanders who have completed training on both the Computer Assisted Instruction lessons and SIMCAT prior to taking the STX in the tank. Please make estimates of performance (percent of "Go's") on the Tank Commanders STX, under each of the following eight sets of conditions.

INSTRUCTOR	STUDENT	TASK	PERCENT OF "GO's" ON STX (TECIT I)
13. Easy	Fast	Easy	
14. Easy	Fast	Tough	
15. Easy	Slow	Easy	
16. Tough	Fast	Easy	
17. Easy	Slow	Tough	
18. Tough	Fast	Tough	
19. Tough	Slow	Easy	
20. Tough	Slow	Tough	

21. Now, please rank the following variables for their importance to the estimations you just made:

Rank	Variable
	Instructors
	Students
	Tasks

Administrator: Sum the eight sets of trials recorded above and divide by 8. Insert this mean value (rounded to a whole number) following the symbol "*N*" in questions 22-24 (TECIT II)

22. If an average student achieves *N* "Go's", how many "Go's" will

... a fast learner receive?
... a slow learner receive?

23. If an average instructor gives *N* "Go's" in training students, how many "Go's" will

... an easy instructor give?
... a tough instructor give?

24. If an average task receives *N* "Go's", how many "Go's" would

... an easy task give?
... a tough task give?

III. Third, think of a group of student tank commanders who <u>have completed training only on the Computer Assisted Instruction lessons prior to taking the STX in the tank.</u> Please make estimates of performance (percent of "Go's") on the Tank Commanders' STX, under each of the following eight sets of conditions.

INSTRUCTOR	STUDENT	TASK	PERCENT OF "GO's" ON STX (TECIT I)
25. Easy	Fast	Easy	
26. Easy	Fast	Tough	
27. Easy	Slow	Easy	
28. Tough	Fast	Easy	
29. Easy	Slow	Tough	
30. Tough	Fast	Tough	
31. Tough	Slow	Easy	
32. Tough	Slow	Tough	

33. Now, please rank the following variables for their importance to the estimations you just made:

Rank	Variable
	Instructors
	Students
	Tasks

Administrator: Sum the eight sets of trials recorded above and divide by 8. Insert this mean value (rounded to a whole number) following the symbol "*N*" in questions 34-36 (TECIT II)

34. If an average student achieves *N* "Go's", how many "Go's" will

... a fast learner receive?
... a slow learner receive?

35. If an average instructor gives *N* "Go's" in training students, how many "Go's" will

... an easy instructor give?
... a tough instructor give?

36. If an average task receives *N* "Go's", how many "Go's" would

... an easy task give?
... a tough task give?

IV. Finally, we will answer similar questions for a group of students who <u>have not had SIMCAT or CAI experience</u>.

INSTRUCTOR	STUDENT	TASK	PERCENT OF "GO's" ON STTX (TECIT II)
37. Easy	Fast	Easy	
38. Easy	Fast	Tough	
39. Easy	Slow	Easy	
40. Tough	Fast	Easy	
41. Easy	Slow	Tough	
42. Tough	Fast	Tough	
43. Tough	Slow	Easy	
44. Tough	Slow	Tough	

45. Now, again rank these variables for their order of importance in determining performance.

Rank	Variable
	Instructors
	Students
	Tasks

Administrator: Sum the trials listed in response to questions 37-44 and divide by 8. Enter this rounded value appropriately following the symbol "*M*" in the three questions that follow. (TECIT II)

---

46. If an average student achieves *M* "Go's", how many "Go's" will

    ... a fast learner receive?
    ... a slow learner receive?

47. If an average instructor gives *M* "Go's" in training students, how many will

    ... an easy instructor give?
    ... a tough instructor give?

48. If an average task receives *M* "Go's", how many "Go's" would

    ... an easy task give?
    ... a tough task give?

TECIT III

1-12 Given the information above, estimate the percent of the students who have participated in the SIMCAT excercise for Student Tank Commander's Course you expect to receive a "Go" for each STTX station.

	Title	Estimated Percent "GO"
1.	____________	____________
2.	____________	____________
3.	____________	____________
4.	____________	____________
5.	____________	____________
6.	____________	____________
7.	____________	____________
8.	____________	____________
9.	____________	____________
10.	____________	____________
11.	____________	____________
12.	____________	____________

13. How many "Go's" do you expect an average student who has participated in SIMCAT to achieve on the STTX?

______Average "Go's"

14-25. Now estimate the percent of the students who have not participated in the SIMCAT excercise for Student Tank Commander's you expect to receive a "GO" for each STTX station.

	Title	Estimated Percent "GO"
14.	____________	____________
15.	____________	____________
16.	____________	____________
17.	____________	____________
18.	____________	____________
19.	____________	____________

	Title	Estimated Percent "GO"
20.	________	________
21.	________	________
22.	________	________
23.	________	________
24.	________	________
25.	________	________

26. How many "Go's" do you expect an average student who has not participated in SIMCAT to achieve on the STX?

    ________Average "Go's"

FORECASTING TRAINING EFFECTIVENESS (FORTE)

OVERVIEW: This questionnaire is designed for senior officers, flight instructors, and experienced squadron pilots in Navy Fleet replacement squadrons.

It elicits information that will enable evaluators to guide the design and execution of transfer of training studies involving flight simulators. We are particularly interested in your estimates of the number of trials a student pilot needs to demonstrate NATOPS-level mastery of a variety of training tasks taught by a variety of instructors both with and without the aid of a flight simulator.

I. First, think of a group of student pilots in your squadron who have completed simulator training prior to checking out in the aircraft. Please make estimates of the number of trials needed for mastery under each of the following eight sets of conditions.

INSTRUCTOR	STUDENT	TASK	NUMBER TRIALS IN AIRCRAFT (FORTE I)
1. Easy	Fast	Easy	
2. Easy	Fast	Tough	
3. Easy	Slow	Easy	
4. Tough	Fast	Easy	
5. Easy	Slow	Tough	
6. Tough	Fast	Tough	
7. Tough	Slow	Easy	
8. Tough	Slow	Tough	

9. Now, please rank the following variables for their importance to the estimations you just made:

Rank	Variable
	Instructors
	Students
	Tasks

Administrator: Sum the eight sets of trials recorded above and divide by 8. Insert this mean value (rounded to a whole number) following the symbol "*N*" in questions 10-12. (FORTE II)

10. If an average student requires *N* trials to learn to mastery, how many trials will

... a fast learner require?
... a slow learner require?

11. If an average instructor requires *N* trials to train students, how many trials will

... an easy instructor need?
... a tough instructor need?

12. If *N* trials are needed for average tasks, how many trials would

... an easy task require?
... a tough task require?

II. Now we will answer similar questions for a group of students who <u>have not had simulator experience</u>.

INSTRUCTOR	STUDENT	TASK	NUMBER TRIALS IN AIRCRAFT (FORTE I)
13. Easy	Fast	Easy	
14. Easy	Fast	Tough	
15. Easy	Slow	Easy	
16. Tough	Fast	Easy	
17. Easy	Slow	Tough	
18. Tough	Fast	Tough	
19. Tough	Slow	Easy	
20. Tough	Slow	Tough	

21. Now, again rank these variables for their order of importance in determining trials to mastery:

Rank	Variable
	Instructors
	Students
	Tasks

Administrator: Sum the trials listed in response to questions 13-20 and divide by 8. Enter this rounded value appropriately following the symbol "*M*" in the three questions that follow. (FORTE II)

---

22. If an average student requires *M* trials-to-mastery, how many trials will

    ... a fast learner need?
    ... a slow learner need?

23. If an average instructor requires *M* trials to train students, how many will

    ... an easy instructor need?
    ... a tough instructor need?

24. If *M* trials are needed for average tasks, how many trials would

    ... an easy task require?
    ... a tough task require?

Appendix C

Definitions and Abbreviations

-Accuracy of estimation - the discrepancy between estimates and "true" or parametric values. Measured in terms of absolute values or statistical standard errors of estimate. For TD/S, primary interest is in analytic and empirical measures of acquisition learning on the TD/S and transfer of training.

-Acquisition learning - refers to initial learning on a TD/S as opposed to relearning, retention or maintenance of skills. Measured in terms of time and performance on a TD/S.

-Acquisition or procurement process - the steps involved in purchasing training, training devices, simulators, weapon systems or other items relevant to the Army.

-Analytic methods - those methods employing definitions, judgments, experience, logic, systems analysis and other non-empirical methods.

-Baseline data and information - those historical methods that employ databases, similar cases, predecessor cases, research literature and meta-analysis to extrapolate from past research and practice to the design and development.

-Bias of estimates - The extent to which analytic or empirical methods consistently overestimate or underestimate "true" or parametric values.

-Confounded measurements - measurements which cannot be clearly attributed to one of several treatments or causes.

-Courseware vs. hardware and software - Courseware is the instructional materials and content, hardware the physical carrier and software the computer programs or electro-mechanical codes or instructions which aid in operating a TD/S, WS or other technology.

-Cross-sectional study design - a design that makes contrasts or comparisons at a fixed point in time or during a given phase of TD/S development. In contrast, longitudinal study designs are conducted over time or TD/S phases by follow-up or follow-back.

-Empirical data and methods - refers to data from direct measurements of the performance of TD/S and/

or the trainees and instructors using these TD/S. For TD/S it includes the measurement of acquisition learning, the transfer of training experiment, reliability/maintainability, utilization and other data.

-Exportability - refers to the potential use and application of a TD/S or training packages to other Army applications beyond the first application for which it was designed.

-Fidelity, physical - the perceived similarity in its static state of a TD/S and the WS(s) for which it was designed.

-Fidelity, functional - the dynamic response characteristics of a simulator, e.g., whether the simulator banks as fast in response to a pilot's aileron control motions as an aircraft would.

-Instructional management - the process of managing the implementation of a TD/S or other training technology. A set of variables hypothesized to be related to the utilization and technology transfer of a TD/S or other training technology.

-Judgmental variances - a statistical method for extracting variance estimates from judgments including variances appropriate to a TD/S such as student variance, task variance, criterion variance, team variance, and error variance. A method to aid in predicting empirical data from judgments.

-Life cycle development phases - the phases through which a TD/S, WS and training program proceed from conception through fielding or implementation. See Chapter 3, Forms 2 and 3 for various phases.

-Longitudinal study design - see cross-sectional study design.

-Masking effects - the extent to which the training effect of a TD/S or other training technology is obscured as a result of other variables.

-Performance measures - measures on a TD/S or a WS which purport to measure relevant performance.

-Reliability of judgmental measurement - the internal consistency of judgments, the extent of agreement among raters; the extent of agreement among raters from one time to another.

-Skill maintenance and retraining schedule. The period of time during which skills decay to a point where it is cost effective to provide formal retrain-

ing or additional practice on-the-job.

-Task analysis - a coherent unit of work or training. May be subdivided as appropriate into subtasks, skills or exercises. In this model, the terms are used generically to denote the level of analysis which may be performed or available at a given time in the TD/S life cycle. When the number of tasks or sub-elements is large, task grouping or task sampling may be advisable for certain analyses.

-Task complexity - the characteristics of tasks that tend to make them more or less difficult to perform, such as the number of steps and sub-steps involved, timing of information input and output and other indicators.

-Task difficulty - the level of difficulty of students performing a task measured by time and performance.

-Task severability - a task, sub-task, or skill which can be taught separately from other tasks. Sequence, prerequisites or enabling objectives are not a concern for the task, sub-task or skill in question.

-TD/S specific variance - the variance (e.g., characteristics and learning processes) which does not generalize or transfer to learning on the WS.

-Transfer of training - in concept, the extent to which learning from a TD/S or course unit generalizes or transfers to learning on a WS or the job. The empirical transfer of the training paradigm is limited to measuring time savings and/or performance improvement for safe tasks, usually applied within the framework of the course rather than the job itself.

-Treatments - in an empirical experiment, the characteristics of experimental and control groups.

Abbreviations

A	Acquisition
AE	Acquisition Efficiency
BNCOC	Basic Non-Commissioned Officers Course
C	Control Group
CAI	Computer-Aided Instruction
CBP	Comparison-Based Prediction
CTEA	Cost and Training Effectiveness Analysis
DEFT	Device Effectiveness Forecasting Technique
E	Experimental Group
FORTE	Forecasting Training Effectiveness
JR	Job Readiness
MAUM	Multi-attribute Utility Assessment Method
MOTNLY	Motion only
NVSMOT	No visual-no motion
OCR	Operating Cost Ratio
PTC	Percent Transfer to Criterion
PT	Percent Transfer
PTM	Percent Transfer Maximum
PTS	Percent Time Saved
PTTS/A	Percent Total Training Time Saved/Added
SIMCAT	Simulation in Combined Arms Training
SME	Subject Matter Expert
STTX	Situational Tactical Training Exercise
T	Transfer
TC	Tank Commanders
TD/S	Training Device/Simulator
TECIT	Training Effectiveness and Cost Interactive Technique
TER	Transfer Effectiveness Ratio
ToT	Transfer of Training
TP	Training Problem
TRP	Transfer Problem Analysis
TT	Transfer Efficiency Analysis
TTFA	Training Technology Field Activities
UR	Utilization Ratio
VISMOT	Visual and Motion
VISNLY	Visual only
WS	Weapon System